

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2006, 13:32:22 ; Search time 4710 Seconds
(without alignments)
11187.660 Million cell updates/sec

Title: US-10-613-990A-31
Perfect score: 927
Sequence: 1 gtcattgtacgcgatcatg.....gtaactttttactgggttt 927

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	927	100.0	927	6	CQ760542 Sequence
2	927	100.0	39675	1	Continuation (47 o
3	927	100.0	338534	1	EC00093 Escherich
4	920.6	99.3	28445	1	Continuation (56 o
5	920.6	99.3	98450	1	Continuation (55 o
6	912.6	98.4	86898	1	AE016772 Escherich
7	912	98.4	1468	1	M10044 E.coli dye
8	910.2	98.2	107203	1	Continuation (46 o
9	910.2	98.2	225944	1	AE016993 Shigella
10	715.6	77.2	88037	1	AE027284 Salmonell
11	715.6	77.2	284233	1	AE016849 Salmonell
12	714	77.0	13852	1	AE008916 Salmonell
13	714	77.0	55700	1	Continuation (48 o
14	712.4	76.9	85229	1	Continuation (46 o
15	615.6	66.4	622	1	AY832936 Escherich
16	612.6	66.1	619	1	AY832937 Escherich
17	610.8	65.9	614	1	AY832938 Escherich
18	610.8	65.9	614	1	AY832940 Escherich

19	607.2	65.5	612	1	AY832939	Escherich
20	577.2	62.3	726	6	AR385083	Sequence
21	562.4	60.7	564	1	AP267619	Escherich
22	562.4	60.7	564	1	AP267620	Escherich
23	562.4	60.7	564	1	AP267621	Escherich
24	560.8	60.5	564	1	AP267627	Escherich
25	560.8	60.5	564	1	AP267628	Escherich
26	560.8	60.5	564	1	AP267629	Escherich
27	560.8	60.5	564	1	AP267631	Escherich
28	560.8	60.5	564	1	AP267632	Escherich
29	560.8	60.5	564	1	AP267633	Escherich
30	560.8	60.5	564	1	AP267635	Escherich
31	560.8	60.5	564	1	AP267637	Escherich
32	559.2	60.3	564	1	AP267622	Escherich
33	559.2	60.3	564	1	AP267623	Escherich
34	559.2	60.3	564	1	AP267624	Escherich
35	559.2	60.3	564	1	AP267625	Escherich
36	559.2	60.3	564	1	AP267626	Escherich
37	559.2	60.3	564	1	AP267630	Escherich
38	559.2	60.3	564	1	AP267636	Escherich
39	559.2	60.3	564	1	AP267638	Escherich
40	557.6	60.2	564	1	AP267634	Escherich
41	549	59.2	549	1	AY698108	Escherich
42	547.8	59.1	551	1	AY698120	Shigella
43	546.8	59.0	550	1	AY698088	Shigella
44	546.8	59.0	550	1	AY698102	Shigella
45	545.8	58.9	549	1	AY698114	Escherich

ALIGNMENTS

RESULT 1
CQ760542
LOCUS CQ760542 927 bp DNA linear PAT 03-MAR-2004
DEFINITION Sequence 31 from Patent EP1382686.
ACCESSION CQ760542
VERSION CQ760542.1 GI:44904047
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Ishikawa,Y., Imaizumi,A., Mateui,K. and Kojima,H.
TITLE Method for producing target substance by fermentation using a bacterial strain lacking the Arca gene
JOURNAL Patent: EP 1382686-A 31 21-JAN-2004;
Ajinomoto Co., Inc. (JP)
FEATURES
Location/Qualifiers
1..927
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
101..817
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAP32793.1"
/db_xref="GI:44904048"
/translation="MQTPHILIVDELVRLTLKSIPEAGYDVEATDGAEMHQILS
EYDNLVIMDINLPGKGLLARELREQANVALMPLTGRDNEVDKILGLETCADGYIT
KPFNPRLTIRARLLSRNLTGTVSERRSVSYKFGWELDINSRLIGDGRQYK
LPREFRMLHFCEINPGKIQSRAELLKQMTGRELKPHDRTVDVTIRIRKHFESTPTD
PELIATIHGEGRYFCGDL"

ORIGIN
Query Match 100.0%; Score 927; DB 6; Length 927;
Best Local Similarity 100.0%; Pred. No. 3.8e-262;
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATGTTACCCGATCATGTTAATTTGCAGTCATCAGCGTTCAGGACTTTGCT 60
|||||

Db	1	GTCTGTTTACGCCGATCATGTTAAATTTTCAGCATGATCAGCAGGTTCAGGACTTTTGT	60
Qy	61	ACTTCTGTTTCGATTTAGTTGGCAATTTAGTAGCAAAACATGCAAGACCCCGCACATTCT	120
Db	61	ACTTCTGTTTCGATTTAGTTGGCAATTTAGTAGCAAAACATGCAAGACCCCGCACATTCT	120
Qy	121	TATCGTTTGAAGACGAGTTGTTGAACACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGG	180
Db	121	TATCGTTTGAAGACGAGTTGTTGAACACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGG	180
Qy	181	CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCAATCAGATCCTCTCTGAATATGA	240
Db	181	CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCAATCAGATCCTCTCTGAATATGA	240
Qy	241	CATCAACCTGGTGATCATGGATATCAATCTGCGGGTGAAGACGGTCTCTGTAGCGGG	300
Db	241	CATCAACCTGGTGATCATGGATATCAATCTGCGGGTGAAGACGGTCTCTGTAGCGGG	300
Qy	301	TGAACCTGCGGAGCAGGCGAATGTTGCGTTGATGTTCTGACTGCGCGTGACCAAGAAAT	360
Db	301	TGAACCTGCGGAGCAGGCGAATGTTGCGTTGATGTTCTGACTGCGCGTGACCAAGAAAT	360
Qy	361	CGATAAAATTTCTCGGCTCGAAATCGGTGCAGATGACTACATCACCACCAACCGTTCAACCC	420
Db	361	CGATAAAATTTCTCGGCTCGAAATCGGTGCAGATGACTACATCACCACCAACCGTTCAACCC	420
Qy	421	GGTGAACTGACGATTCGTGCGAGGCAACCTACTGTCCCGTACATGAATCTGGGTACTGT	480
Db	421	GGTGAACTGACGATTCGTGCGAGGCAACCTACTGTCCCGTACATGAATCTGGGTACTGT	480
Qy	481	CAGCGAAGACGTCGTGAGGTTGAAAGCTACAAGTTCAATGTTGGAAGTGGACATCAA	540
Db	481	CAGCGAAGACGTCGTGAGGTTGAAAGCTACAAGTTCAATGTTGGAAGTGGACATCAA	540
Qy	541	CAGCGGTTCTGTTGATCGGCGCTGATGGCGAGCAGTACAAGCTGCGCGCAGCGAGTTCGG	600
Db	541	CAGCGGTTCTGTTGATCGGCGCTGATGGCGAGCAGTACAAGCTGCGCGCAGCGAGTTCGG	600
Qy	601	CGCATGCTTCACTTCTGTGAAAACCCAGCGAAAATTCAGTCCCGTGTGAACTGCTGAA	660
Db	601	CGCATGCTTCACTTCTGTGAAAACCCAGCGAAAATTCAGTCCCGTGTGAACTGCTGAA	660
Qy	661	GAAATGACCGCGTGACGTGAAACCGCACCGTACTGCTAGACGTGACGATCCGCG	720
Db	661	GAAATGACCGCGTGACGTGAAACCGCACCGTACTGCTAGACGTGACGATCCGCG	720
Qy	721	TATTCGTAAACATTTTTCGAAATCTACGCGGATACGCGGAAATTCATCGCCACCATTCACGG	780
Db	721	TATTCGTAAACATTTTCGAAATCTACGCGGATACGCGGAAATTCATCGCCACCATTCACGG	780
Qy	781	TGAAGGTTATCGTTCTGCGGTGATCTGGAAGATTAATCGGCTTTTACGACCGTCAAAAA	840
Db	781	TGAAGGTTATCGTTCTGCGGTGATCTGGAAGATTAATCGGCTTTTACGACCGTCAAAAA	840
Qy	841	AACGGGCTTTTTCAGCGCGTTTATTTTCAACCTTATTTCCAGATACGTAACCTCATC	900
Db	841	AACGGGCTTTTTCAGCGCGTTTATTTTCAACCTTATTTCCAGATACGTAACCTCATC	900
Qy	901	GTCCGTTGTAACCTCTTTTACTGCGTTT	927
Db	901	GTCCGTTGTAACCTCTTTTACTGCGTTT	927

RESULT 2
U00096_46/c
WPCOMMENT
Sequence split into 47 fragments LOCUS U00096 Accession U00096
Fragment Name Begin End
U00096_00 1 110000
U00096_01 100001 210000
U00096_02 200001 310000
U00096_03 300001 410000
U00096_04 400001 510000

U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000
U00096_26	2600001	2710000
U00096_27	2700001	2810000
U00096_28	2800001	2910000
U00096_29	2900001	3010000
U00096_30	3000001	3110000
U00096_31	3100001	3210000
U00096_32	3200001	3310000
U00096_33	3300001	3410000
U00096_34	3400001	3510000
U00096_35	3500001	3610000
U00096_36	3600001	3710000
U00096_37	3700001	3810000
U00096_38	3800001	3910000
U00096_39	3900001	4010000
U00096_40	4000001	4110000
U00096_41	4100001	4210000
U00096_42	4200001	4310000
U00096_43	4300001	4410000
U00096_44	4400001	4510000
U00096_45	4500001	4610000
U00096_46	4600001	4639675

Continuation (47 of 47) of U00096 from base 4600001 (U00096 Escherichia coli K-12)

Query Match	100.0%;	Score	927;	DB	1;	Length	39675;		
Best Local Similarity	100.0%;	Pred. No.	7.9e-262;						
Matches	927;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;

Qy	1	GTCTGTTTACGCCGATCATGTTAAATTTGCAGCATGCATCAGGACAGGTTCAGGACTTTTGT	60
Db	38429	GTCTGTTTACGCCGATCATGTTAAATTTGCAGCATGCATCAGGACAGGTTCAGGACTTTTGT	38370
Qy	61	ACTTCTGTTTCGATTTAGTTGGCAATTTAGTAGCAAAACATGCAGACCCCGCACATTCT	120
Db	38369	ACTTCTGTTTCGATTTAGTTGGCAATTTAGTAGCAAAACATGCAGACCCCGCACATTCT	38310
Qy	121	TATCGTTGAAGACGAGTTGTTAAACACGCAACAGTTTGAAAAGTATTTTCGAAGCGGAAGG	180
Db	38309	TATCGTTGAAGACGAGTTGTTAAACACGCAACAGTTTGAAAAGTATTTTCGAAGCGGAAGG	38250
Qy	181	CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCATCAGATCCTCTCTGAATATGA	240
Db	38249	CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCATCAGATCCTCTCTGAATATGA	38190
Qy	241	CATCAACTGTGTGATCATGGATATCAATCTGCGGGTGAAGAACGGTCTTCTGTAGCGCG	300
Db	38189	CATCAACTGTGTGATCATGGATATCAATCTGCGGGTGAAGAACGGTCTTCTGTAGCGCG	38130
Qy	301	TGAACCTGCGGAGCAGGCGAATGTTGGTTGATGTTCTTGACTGCGCGTGACCAACGAAT	360
Db	38129	TGAACCTGCGGAGCAGGCGAATGTTGGTTGATGTTCTTGACTGCGCGTGACCAACGAAT	38070
Qy	361	CGATAAAATTTCTCGGCTTCGAATCGGTGCAGATGACTACATCACCAACCGTTCAACCC	420

Db 38069 CGATAAAATTCCTCGGCTCGAAATCGGTGCAGATGACTACATCACCAACCGTTCAACCC 38010
QY 421 GCGTGAACAGCATTCGTCGACGCAACTACTGTCCTGACCATGATTCGGGTACTGT 480
Db 38009 GCGTGAACAGCATTCGTCGACGCAACTACTGTCCTGACCATGATTCGGGTACTGT 37950
QY 481 CAGCGAAGAACCTGCTAGCGTTGAAAGCTCAAGTTCATGTTGGAATCGACATCAA 540
Db 37949 CAGCGAAGAACCTGCTAGCGTTGAAAGCTCAAGTTCATGTTGGAATCGACATCAA 37890
QY 541 CAGCGTTGCTGATCGGCTGATGCGAGCAGTCAAGTTCGCGCGGAGGAGTTCG 600
Db 37889 CAGCGTTGCTGATCGGCTGATGCGAGCAGTCAAGTTCGCGCGGAGGAGTTCG 37830
QY 601 CCGCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCGCTGCTGAATCTGCTGAA 660
Db 37829 CCGCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCGCTGCTGAATCTGCTGAA 37770
QY 661 GAAATGACCGCGCTGAGTGAACCCGACGACCGTACTGTGACGTGACGATCCGCG 720
Db 37769 GAAATGACCGCGCTGAGTGAACCCGACGACCGTACTGTGACGTGACGATCCGCG 37710
QY 721 TATTCGTAACATTTGCAATCTACGCGGATACGCGGGAATCATCGCACCATTCACGG 780
Db 37709 TATTCGTAACATTTGCAATCTACGCGGATACGCGGGAATCATCGCACCATTCACGG 37650
QY 781 TCAAGGTTATCGTCTCGGGTGATCTGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 840
Db 37649 TCAAGGTTATCGTCTCGGGTGATCTGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 37590
QY 841 AACGGCGCTTTTAGCGCGTTTTATTTTTCAACCTTATTTCCAGATACGTAATCATC 900
Db 37589 AACGGCGCTTTTAGCGCGTTTTATTTTTCAACCTTATTTCCAGATACGTAATCATC 37530
QY 901 GTCGTTGTAATCTTCTACTGCTTT 927
Db 37529 GTCGTTGTAATCTTCTACTGCTTT 37503

RESULT 3
ECOW93/c
LOCUS Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.
DEFINITION U14003
ACCESSION U14003.1 GI:1263172
VERSION U14003.1
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 338534)
AUTHORS Burland,V., Plunkett,G. III, Sofia,H.J., Daniels,D.L. and Blattner,F.R.
TITLE Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes
JOURNAL Nucleic Acids Res. 23 (12), 2105-2119 (1995)
PUBMED 7610040
REFERENCE 2 (bases 1 to 338534)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
COMMENT On Apr 17, 1996 this sequence version replaced gi:536929.
This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. The start of this entry overlaps the end of the entry ECOW89 (U00006) by 1885 bp.

FEATURES
source

Location/Qualifiers
1. 338534
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/map="92.8 to 100 minutes"
/note="This sequence comprises the following lambda clones: DD928(EC27-278), DD930(EC30-188), DD933 (EC27-1320), DD935 (EC30MM32), DD937 (EC27-1070), DD941 (EC17-142), DD945 (EC17-8), DD947 (EC44A-34), DD949 (EC22-169), DD952 (EC27-297), DD953 (EC27-409), DD956 (EC17-103), DD958 (EC30MM1), DD960 (EC30MM2), DD962 (EC21-104), DD965 (EC23A-40), DD968 (EC30K660A-4pp), DD970 (EC19-202), DD974 (EC19-61), DD975 (EC18-233), DD977 (EC30MM7), DD980 (EC27-1151), DD981 (EC18-3), DD984 (EC17-101), DD987 (EC18-115), DD990 (EC17-136), DD992 (EC18-282), DD995 (EC17-115), DD997 (EC18-113); M13mp19 or Janus vectors were used for subcloning"
1. 1885
/note="1885 bp overlap with end of GenBank Accession Number U00006 (ECOW89)"
CDS
complement (<1. 398)
/note="ORF f326 of GenBank Accession Number U00006 (ECOW89)"
/codon_start=1
/transl_table=11
/label_ORF_f132p
/protein_id="AAA96985.1"
/db_xref="GI:1263173"
/translation="MGFTTRVKESEBKKPFNFALFMDKYGTFFILAIIVAFGSLSP EYFTTNNTIOIFVQSSVTVLIGMEFFAILVAGIDLSVGAILALSGMVTAKMLAGV DPFLAAMTGGVLVGGALGANGCLVNMTEL"
277. 397
/note="predicted bend of 75.10 degrees"
complement (377. 1909)
/note="ORF f510"
/codon_start=1
/transl_table=11
/label_ORF_f510
/protein_id="AAA96986.1"
/db_xref="GI:536931"
/translation="MATPYISMAGIKSGFGVHALKSVNLTVYFGEIHALLGENGAGK STLKMLSGIHEPTKTTITNNISYKLDHLKLAQGLIGIYQELSDIETLVLENLY IGRHLTKICGVNIIDWREMRVRAAMLLRVGLKVDLDEKVANLSISHKOMLEIAKTL MLDKAVIMDEPTSSLTKNEYDYLFLIMNQLRKEGTALVYISHKLAERRRICDRVTVM KDSSVCSGIVSDNDDIVRLMVGRLOLNFNKNVSNLAHETVFEVRNVSRRD KKVREDIIFSVCGRGILFAGLVGSGRTLMNCLFGVDKRAGEIRLNGKDISPRSPLD AVKKGMAVITSERRDNGFFPNFSAQNMALSRSLKDGKYGKMGILFHEVDQRTAENQ RELALAKHSVNQNTLSLGGNQOKVLISKWLCCEPVIIFDEPTRGIDVGAKAIYK VMRQLADGKVLVMSSELPEIITVCDRIAVFCSEGLTQILTRDDMSSEIMAWALP QR"
809. 931
/note="predicted bend of 74.75 degrees"
1768. 1889
/note="predicted bend of 75.87 degrees"
complement (1983. 2010)
/note="promoter matrix score of 55; putative; resembles flh-regulated promoter"
complement (1985. 2021)
/standard_name="REP; repetitive extragenic palindromic element"
/note="contains 1 REP sequence"
complement (2036. 2971)
/note="ORF_f311"
/codon_start=1
/transl_table=11
/label_ORF_f311
/protein_id="AAA96987.1"
/db_xref="GI:536932"
/translation="MNKYLKYPFGTLVGLMLSTAFAPAAAYAVLVKLTISNPFVDMKK GIEDEAKTLGVSVDIFAPSPSEGDFQSQQLFEDLSNKNYKGIAPALSSVNLVMPVAR

AKKGIYLVNLDKIDMONLKKAGGNVEAFVTTDNVAVGAKGASFIIDKLGAEGEVA
IEGKAGNAGEARRNGATEAFKQASQIKLVASQPADMDRIKALDVAATNVLQRPNIK
AIYCHADWAMGVAQAVANAGTKVLVVGTDGIPKARKMVEAQMTATVAQNPAIDIG
ATGLKMDYDAKSGKVIPLDKAPEFKLVDSILVTQ"

misc_structure

3010. .3121
/note="predicted bend of 73.22 degrees"
complement (3030. .3953)

CDS

/note="ORF_f307"
/codon_start=1
/transl_table=11
/label=ORF_f307

/protein_id="AAA96988.1"
/db_xref="GI:536933"

/translation="WASSIQOENVMQSSEFDSALPNGIGLAPYLKMQEGWENESR
IVELKPGNLSCAPAINABALAVSEAMIVKSKLIGFSGFNLRSALSDYFSQSE
QVLPSELADPAQDVKNVFNITLRTIMEGQSVINVDIHRAAFFQQRORDLYGA
GGSNAICADVOKFIKVRQVAPDAIHMMMSASLQEGDVLVVTHSGRTSDVKAA
VELAKNGAKIICITHSVHSPIAKLADYIICSPAPETLLGRNASARILQLTLLDAPP
VSVQNLTEQANINMOKTGAVDFFSPGALK"

promoter

complement (3952. .3979)
/note="promoter matrix score of 61; putative"
4151. .4251

misc_structure

/note="predicted bend of 72.06 degrees"
4279. .4728

gene

/gene="yjca"

CDS

4279. .4728
/gene="yjca"
/codon_start=1
/transl_table=11
/label=ORF_0149
/protein_id="AAA96989.1"
/db_xref="GI:536934"

/translation="MKKIFAGCDHVGFIKHEIVAHVVERGVVEIDKGTWMSERTDYP
HYASQVALAVGEVDGILICGTGVGISIAAKFAGIYAVVVCSEPSYQALSQRHNDT
NVAFSGRVGLGELAKMTVDWLGAQYEGSRHQQRVEAITAEQRN"

misc_feature

complement (4520. .16191)
/note="corresponds to GenBank Accession Number D90227
(ECOPHN) 1. .11677; the related sequence GenBank Accession
Number J05260 (ECOPHNAQ) is from E. coli B, not K-12"

gene

complement (4739. .5146)
/gene="phnq"

CDS

complement (4739. .5146)
/gene="phnq"
/note="annotated in GenBank Accession Number J05260
(ECOPHNAQ), but probably not a gene"
/codon_start=1
/transl_table=11

/protein_id="AAA96990.1"
/db_xref="GI:536935"
/translation="WPAPYPTSSPILIFHSPILFIQLPILTLIKLALAIQVLPAAIFKL
AAVAVILSAPIVAHFTLPAVAVARIFVSTVRVIVSINRLNLIANGVVDNIGLGRND
SVSRVKNRNSGRKFISQPKETLSLIITHPCK"

gene

complement (5272. .6030)
/gene="phnp"

CDS

complement (5272. .6030)
/gene="phnp"
/codon_start=1
/transl_table=11
/protein_id="AAA96991.1"
/db_xref="GI:536936"

/translation="WSLITLITLGTGAGQVPWAGCECAAPARSPQVRRPQSGVV
KNDALITLIDAGHLADRWSFGSQQLLTHYHDHVQGLFPLRWGVDPIPYVGP
DSQGDLDLPKGLLDIFSHVFPFVVDLQGLQVTPPLNHSKLTFGYLLETAFHSRVA
WLSDTAGLPKTKFLRNQPVQVMVDCSHPPRADA PRNHCDLNTVLALNQVIRSPV
ILTHISHQPDAMLMENALPSGEVFGDGMIEGVA"

misc_structure

5452. .5586
/note="predicted bend of 75.60 degrees"
complement (6032. .6466)

gene

/gene="phnq"

CDS

complement (6032. .6466)
/gene="phnq"
/codon_start=1
/transl_table=11

/protein_id="AAA96992.1"
/db_xref="GI:536937"
/translation="MPACELRPATQYDITDAVYALI CELKQEPDHPHAFPRVGFENALRD
PNRYHIALDLGEVGMIGLHLQPHLHVNMWIGIBELVVMFQARGLNVGSKLLAWAE
EEARQAGAEMTELSITNKRHDHRFYLRGEYEQSHFRFTKAL"
complement (6453. .7010)
/gene="phnN"
complement (6453. .7010)
/gene="phnN"
/note="HisP-like nucleotide binding protein (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/protein_id="AAA96993.1"
/db_xref="GI:536938"
/translation="MMOKLIWMPGSSGKDSLAEI RLREQTOLLVHRYITRDASA
GSENHAIQSQEFTRAGQNLALSHWANGLYGVGEIDLMHAGFDVLVNGSRAHL
POARARYQSALLPVCLQVSPFELRQLENRGRENASEINARLARAARYTPQDCHTLNN
DGSUROSVDTLTLIHQKEKHACL"
complement (7010. .8146)
/gene="phnM"
complement (7010. .8146)
/gene="phnM"

Query Match 100.0%; Score 927; DB 1; Length 338534;

Best Local Similarity 100.0%; Pred. No. 1.2e-261; Mismatches 0; Indels 0; Gaps 0;
Matches 927; Conservative 0; TGCAGCATGTCATCAGGAGTTCAGGAGTTCCTTGT 331180

QY	1	GTCATGTTACGCCGATCATGTTAAATTTGCAGCATGTCATCAGGAGTTCAGGAGTTCCTTGT	60
DB	331239	GTCAATGTTACGCCGATCATGTTAAATTTGCAGCATGTCATCAGGAGTTCAGGAGTTCCTTGT	331180
QY	61	ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAACATGCAGACCCCGCACATTC	120
DB	331179	ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAACATGCAGACCCCGCACATTC	331120
QY	121	TATCGTTGAACACAGTTGTGTAACACGCAACAGTGTGAAAGATATTTTCGAAGCGGAAG	180
DB	331119	TATCGTTGAACACAGTTGTGTAACACGCAACAGTGTGAAAGATATTTTCGAAGCGGAAG	331060
QY	181	CTATGATGTTTTGAAAGCGACAGATGGCGCGGAATATGCATCAGATTCCTCTCTGAATATGA	240
DB	331059	CTATGATGTTTTGAAAGCGACAGATGGCGCGGAATATGCATCAGATTCCTCTCTGAATATGA	331000
QY	241	CATCAACCTGTTGATCATGGATATCAATCTCCCGGTTAAGAACGCTCTCTGTTAGCGCG	300
DB	330999	CATCAACCTGTTGATCATGGATATCAATCTCCCGGTTAAGAACGCTCTCTGTTAGCGCG	330940
QY	301	TGAACCTGCGCAGCAGGCGGAATGTTGCGTGTGATGTTCTCTGACTGCGCGTGACACGAAAT	360
DB	330939	TGAACCTGCGCAGCAGGCGGAATGTTGCGTGTGATGTTCTCTGACTGCGCGTGACACGAAAT	330880
QY	361	CGATAAAATTTCTCGGCTTCGAAATCGGTGCGAGATGACTACATCAACAAACCGTTCAACCC	420
DB	330879	CGATAAAATTTCTCGGCTTCGAAATCGGTGCGAGATGACTACATCAACAAACCGTTCAACCC	330820
QY	421	CGGTGAACCTGAGGATTCGTGACGCAACCTACTGTCCTGCTGCTGCTGCTGCTGCTGCTG	480
DB	330819	CGGTGAACCTGAGGATTCGTGACGCAACCTACTGTCCTGCTGCTGCTGCTGCTGCTGCTG	330760
QY	481	CAGCGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
DB	330759	CAGCGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	330700
QY	541	CAGCGCTTCTGTTGATCGGCTTCGATGGCGAGCAGTACAAGCTGCGCGCAGCGAGTTCGG	600
DB	330699	CAGCGCTTCTGTTGATCGGCTTCGATGGCGAGCAGTACAAGCTGCGCGCAGCGAGTTCGG	330640
QY	601	CGCCATGCTTCACTTCTGTGAAAACCCAGGAAATTCAGTCCCGTCTGCTGCTGCTGCTG	660
DB	330639	CGCCATGCTTCACTTCTGTGAAAACCCAGGAAATTCAGTCCCGTCTGCTGCTGCTGCTG	330580
QY	661	GAATAATGACCGGCGCTGAGCTGAAACCGCACGACCGTACTGTAGACGTCGATCCGCGCG	720

Db 330579 GAAATGACCGCGGTGAGCTGAAACCGCACGACCGTACTGTAGACGTGACGATCCGCG 330520
Qy 721 TATTGCTAAACATTTTGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 780
Db 330519 TATTGCTAAACATTTTGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 330460
Qy 781 TGAAGTTATCGCTTCTGCGGTGATCTGGAAGATTAAATCGGCTTTTACCAACCGTCAAAAA 840
Db 330459 TGAAGTTATCGCTTCTGCGGTGATCTGGAAGATTAAATCGGCTTTTACCAACCGTCAAAAA 330400
Qy 841 AACGGGCTTTTACGCGGTTTATTTTCAACCTTATTTCCAGATACGTAACCTCATC 900
Db 330399 AACGGGCTTTTACGCGGTTTATTTTCAACCTTATTTTCCAGATACGTAACCTCATC 330340
Qy 901 GTCCGTTGTAACCTTCTTACTGCGCTTT 927
Db 330339 GTCCGTTGTAACCTTCTTACTGCGCTTT 330313

RESULT 4
AE005174_55/c
WPCOMMENT

Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000
AE005174_33	3300001	3410000
AE005174_34	3400001	3510000
AE005174_35	3500001	3610000
AE005174_36	3600001	3710000
AE005174_37	3700001	3810000
AE005174_38	3800001	3910000
AE005174_39	3900001	4010000
AE005174_40	4000001	4110000
AE005174_41	4100001	4210000
AE005174_42	4200001	4310000
AE005174_43	4300001	4410000
AE005174_44	4400001	4510000
AE005174_45	4500001	4610000
AE005174_46	4600001	4710000
AE005174_47	4700001	4810000

AE005174_48 4800001 4910000
AE005174_49 4900001 5010000
AE005174_50 5000001 5110000
AE005174_51 5100001 5210000
AE005174_52 5200001 5310000
AE005174_53 5300001 5410000
AE005174_54 5400001 5510000
AE005174_55 5500001 5528445
Continuation (56 of 56) of AE005174 from base 5500001 (AE005174 Escherichia coli O157:H7)

Query Match 99.3%; Score 920.6; DB 1; Length 28445;
Best Local Similarity 99.6%; Pred. No. 5.7e-260;
Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTCATGTTACGCCCATCATGTTAATTTGCAGCATGTCATCAGGCAGGTTCAGGACCTTTTGT 60
Db 271199 GTCATGTTACGCCCATCATGTTAATTTGCAGCATGTCATCAGGCAGGTTCAGGACCTTTTGT 271140
Qy 61 ACTTCCTGTTTCGATTTAGTTGGCAATTTTAGTTAGCAAAATGCAAGACCCCGCAATTCT 120
Db 271139 ACTTCCTGTTTCGATTTAGTTGGCAATTTTAGTTAGCAAAATGCAAGACCCCGCAATTCT 27080
Qy 121 TATCGTTGAACGACGAGTTGGTAAACGCAACACGTTGAAAGTATTTTCGAGCGGAAG 180
Db 27079 TATCGTTGAACGAGTTGGTAAACGCAACACGTTGAAAGTATTTTCGAGCGGAAG 27020
Qy 181 CTATGATGTTTTCGAAAGCAGATGCGCGGAAATGCATCAGATCCTCTCTGAATATGA 240
Db 27019 CTATGATGTTTTCGAAAGCAGATGCGCGGAAATGCATCAGATCCTCTCTGAATATGA 26960
Qy 241 CATCAACCTGTGATCATGATATCAATCTCCCGGTAAAGAACGGTCTTCTGTTAGCGCG 300
Db 26959 CATCAACCTGTGATCATGATATCAATCTCCCGGTAAAGAACGGTCTTCTGTTAGCGCG 26900
Qy 301 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGCTGCGCTGCAACGAACT 360
Db 26899 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGCTGCGCTGCAACGAACT 26840
Qy 361 CGATAAAATTTCTCGGCCCTCGAAATCGGTGCGAGATGACTACATCAACAAACCGTTCAACCC 420
Db 26839 CGATAAAATTTCTCGGCCCTCGAAATCGGTGCGAGATGACTACATCAACAAACCGTTCAACCC 26780
Qy 421 CGGTGAACCTGACGATTCGTGCAACCTACTGTCCGCTACCATGAAATCTGGGTACTGT 480
Db 26779 CGGTGAACCTGACGATTCGTGCAACCTACTGTCCGCTACCATGAAATCTGGGTACTGT 26720
Qy 481 CAGCGAAGACGTGCTAGCGTTGAAAGCTACAAGTTCAATGGTTGGGAACCTGGACATCAA 540
Db 26719 CAGCGAAGACGTGCTAGCGTTGAAAGCTACAAGTTCAATGGTTGGGAACCTGGACATCAA 26660
Qy 541 CAGCGTTTCGTTGATCGGCCCTCGATGCGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCGG 600
Db 26659 CAGCGTTTCGTTGATCGGCCCTCGATGCGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCGG 26600
Qy 601 CGCCATGTTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA 660
Db 26599 CGCCATGTTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA 26540
Qy 661 GAAATGACCGCGCGTGTGCTGAAACCCGCAACCGTACTGTAGACGTGACGATCCGCG 720
Db 26539 GAAATGACCGCGCGTGTGCTGAAACCCGCAACCGTACTGTAGACGTGACGATCCGCG 26480
Qy 721 TATTGTAACATTTTCGAATCTACGCCGGATACGCCGGAATTCATCGCCACCATTCACGG 780
Db 26479 TATTGTAACATTTTCGAATCTACGCCGGATACGCCGGAATTCATCGCCACCATTCACGG 26420
Qy 781 TGAAGTTTATCGCTTCTCGCGTGTATCTGGAAGATTAAATCGGCTTTTACCGTCAAAAA 840
Db 26419 TGAAGTTTATCGCTTCTCGCGTGTATCTGGAAGATTAAATCGGCTTTTACCGTCAAAAA 26360
Qy 841 AACGGGCTTTTACGCGCGTTTATTTTCAACCTTATTTCCAGATACGTAACCTCATC 900
Db 26359 AACGGGCTTTTACGCGCGTTTATTTTCAACCTTATTTTCCAGATACGTAACCTCATC 26300

QY 901 GTCCGTTGTAACCTCTTTTACTGCGCTTT 927
DB 26299 GTCCGTTGTAACCTCTTTTACTGCGCTTT 26273

RESULT 5

BA000007_54/c

WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name Begin End

BA000007_00 1 110000

BA000007_01 100001 210000

BA000007_02 200001 310000

BA000007_03 300001 410000

BA000007_04 400001 510000

BA000007_05 500001 610000

BA000007_06 600001 710000

BA000007_07 700001 810000

BA000007_08 800001 910000

BA000007_09 900001 1010000

BA000007_10 1000001 1110000

BA000007_11 1100001 1210000

BA000007_12 1200001 1310000

BA000007_13 1300001 1410000

BA000007_14 1400001 1510000

BA000007_15 1500001 1610000

BA000007_16 1600001 1710000

BA000007_17 1700001 1810000

BA000007_18 1800001 1910000

BA000007_19 1900001 2010000

BA000007_20 2000001 2110000

BA000007_21 2100001 2210000

BA000007_22 2200001 2310000

BA000007_23 2300001 2410000

BA000007_24 2400001 2510000

BA000007_25 2500001 2610000

BA000007_26 2600001 2710000

BA000007_27 2700001 2810000

BA000007_28 2800001 2910000

BA000007_29 2900001 3010000

BA000007_30 3000001 3110000

BA000007_31 3100001 3210000

BA000007_32 3200001 3310000

BA000007_33 3300001 3410000

BA000007_34 3400001 3510000

BA000007_35 3500001 3610000

BA000007_36 3600001 3710000

BA000007_37 3700001 3810000

BA000007_38 3800001 3910000

BA000007_39 3900001 4010000

BA000007_40 4000001 4110000

BA000007_41 4100001 4210000

BA000007_42 4200001 4310000

BA000007_43 4300001 4410000

BA000007_44 4400001 4510000

BA000007_45 4500001 4610000

BA000007_46 4600001 4710000

BA000007_47 4700001 4810000

BA000007_48 4800001 4910000

BA000007_49 4900001 5010000

BA000007_50 5000001 5110000

BA000007_51 5100001 5210000

BA000007_52 5200001 5310000

BA000007_53 5300001 5410000

BA000007_54 5400001 5498450

Continuation (55 of 55) of BA000007 from base 5400001 (BA000007 Escherichia coli O157:H7

Query Match 99.3%; Score 920.6; DB 1; Length 98450;

Best Local Similarity 99.6%; Pred. No. 7.3e-260;

Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTCATGTTACGCCGATCATGTTAAATTTGCAGCATGCATCAGGCAGGTCAGGACTTTTGT 60

DB 97204 GTCATGTTACGCCGATCATGTTAAATTTGCAGCATGCATCAGGCAGGTCAGGACTTTTGT 97145
QY 61 ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAACATGACAGCCCGCACATTTCT 120
DB 97144 ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAACATGACAGCCCGCACATTTCT 97085
QY 121 TATCGTTGAAGACAGAGTTGGTAACACGCAACCGTTGAAAAAGTATTTTCGAAGCGGAAGG 180
DB 97084 TATCGTTGAAGACAGAGTTGGTAACACGCAACCGTTGAAAAAGTATTTTCGAAGCGGAAGG 97025
QY 181 CTATGATGTTTTCGAAGACAGATGCGCGGGAATGATCAGATCCCTCTCTGAATATGA 240
DB 97024 CTATGATGTTTTCGAAGACAGATGCGCGGGAATGATCAGATCCCTCTCTGAATATGA 96965
QY 241 CATCAACCTGCTGATCATGATATCAATCTGCCGGGTAAAGAACGGTCTTCTGTTAGCCGG 300
DB 96964 CATCAACCTGCTGATCATGATATCAATCTGCCGGGTAAAGAACGGTCTTCTGTTAGCCGG 96905
QY 301 TGAACCTGCGAGCAGCGCAATGTTGCTTCTGATGTTCTGACTGCGCCGTGACACGAAGT 360
DB 96904 TGAACCTGCGAGCAGCGCAATGTTGCTTCTGATGTTCTGACTGCGCCGTGACACGAAGT 96845
QY 361 CGATAAAATTTCTCGGCCCTCGAAATCGGTGCAAGATGACTACATCAACCAACCGTTCAACCC 420
DB 96844 CGATAAAATTTCTCGGCCCTCGAAATCGGTGCAAGATGACTACATCAACCAACCGTTCAACCC 96785
QY 421 GCGTGAACCTGACGATTCGTGACCGCAACCTACTGTCGGTACCATGATCTCGGTACTGTT 480
DB 96784 GCGTGAACCTGACGATTCGTGACCGCAACCTGTCCTGCGGTACCATGATCTCGGTACTGTT 96725
QY 481 CAGCGAAGACGTCGTAGCGTTGAAAGCTCAAGTTCAATGTTGGGAACTGGACATCAA 540
DB 96724 CAGCGAAGACGTCGTAGCGTTGAAAGCTCAAGTTCAATGTTGGGAACTGGACATCAA 96665
QY 541 CAGCGGTTTCGTTGATCGGCCCTGATGCGGACGACGTACAAGTGGCGGACGAGTTCGG 600
DB 96664 CAGCGGTTTCGTTGATCGGCCCTGATGCGGACGACGTACAAGTGGCGGACGAGTTCGG 96605
QY 601 CGCCATGCTTCACCTCTGTGAAAAACCCAGGCAAAATTCAGTCCGCTGCTGAACCTGCTGAA 660
DB 96604 CGCCATGCTTCACCTCTGTGAAAAACCCAGGCAAAATTCAGTCCGCTGCTGAACCTGCTGAA 96545
QY 661 GAAATGACCGCGCGTGTAGCTGAAACCGCACGACCGTACTGTAGACGTGACGATCCGCGG 720
DB 96544 GAAATGACCGCGCGTGTAGCTGAAACCGCACGACCGTACTGTAGACGTGACGATCCGCGG 96485
QY 721 TATTCGTAACCAATTCGAATCTACCGCGGATACCGCGGAAATCATTCGCCACATTCACGG 780
DB 96484 TATTCGTAACCAATTCGAATCTACCGCGGATACCGCGGAAATCATTCGCCACATTCACGG 96425
QY 781 TGAAGGTTATCGCTTCTCGGTGATCTGGAAGATTAATCGGCTTTACACCGTCAAAAA 840
DB 96424 TGAAGGTTATCGCTTCTCGGTGATCTGGAAGATTAATCGGCTTTACACCGTCAAAAA 96365
QY 841 AACCGCGCTTTTATAGCGCGCTTTTATTTTCAACCTTATTTCCAGATACGTAACATC 900
DB 96364 AACCGCGCTTTTATAGCGCGCTTTTATTTTCAACCTTATTTCCAGATACGTAACATC 96305
QY 901 GTCCGTTGTAACCTTCTTTTACTGGCTTT 927
DB 96304 GTCCGTTGTAACCTTCTTTTACTGGCTTT 96278

RESULT 6

AE016772/c

LOCUS

DEFINITION Escherichia coli CFT073 section 18 of 18 of the complete genome.

ACCESSION AE016772 AE014075

VERSION AE016772.1 GI:26111641

KEYWORDS Escherichia coli CFT073

SOURCE Escherichia coli CFT073

ORGANISM Escherichia coli CFT073

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS

1 (bases 1 to 86898)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157

TITLE

JOURNAL
PUBMEDREFERENCE
AUTHORS

2 (bases 1 to 86898)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. .86898
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(152. .1495)
/gene="gntP"
/locus_tag="c5401"
complement(152. .1495)
/gene="gntP"
/locus_tag="c5401"
/function="transport; Transport of small molecules:
Carbohydrates, organic acids, alcohols"
/note="Glucosylate permease 3; Escherichia coli K-12
ortholog: b4321; Escherichia coli O157:H7 ortholog: z5919"
/codon_start=1
/transl_table=1

FEATURES
source

gene

CDS

CDS

/gene="uxuB"
/locus_tag="c5403"
3100. .4560
/gene="uxuB"
/locus_tag="c5403"
/EC_number="1.1.1.57"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="Escherichia coli K-12 ortholog: b4323; Escherichia
coli O157:H7 ortholog: z5921"
/codon_start=1
/transl_table=11
/product="D-mannanate oxidoreductase"
/protein_id="AAN83825.1"
/db_xref="GI:26111644"
/translation="MTTIVDSNLVAPRSPWDSRLESRIVHLGGCAFRAHQALYTHH
LLESTDSWGICEVNLPGNDRVLIENLKQQLLYTVAKGAESELKLIIGMKALH
PEIDGEGILNMAPQTAIVSLVITKGYCADAAGDLDNNPLIKHLDNPTAPKS
AIGYIVEARLRRLREKGLKFTVNSCDNVRNGHVAKVAVLGLAQARDPQLAAITENV
TFPCTVDRIVPAATPETLQBIADQLGVYDCAICSPFRQWIEDNFVNGRPMWDKV
GAQFVADVVPPEMMLKRLNGSHSLAYLGLVGGVETIADTWTNPAYKAAFPALMQE
QAQTLSPGEGTDLNAVATLLIERPSNLSLHRTWQIAMDGSKLQRLLDVPLHLQN
GGSWRHLALGVAGWRYTQGVDEQNALDVVDPMMAEFQKINAQYGGADRVKALLGLS
GIFADDLPQNADFVGAVTAAAYQQLCERGARECVAAL"
4775. .5548
/gene="uxuR"
/locus_tag="c5404"
4775. .5548
/gene="uxuR"
/locus_tag="c5404"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/note="Escherichia coli K-12 ortholog: b4324; Escherichia
coli O157:H7 ortholog: z5922"
/codon_start=1
/transl_table=11
/product="Uxu operon transcriptional regulator"
/protein_id="AAN83826.1"
/db_xref="GI:26111645"
/translation="MKSATSQRPQVQGVAMIRDLIIKTPVNGERLPPEREIABMLD
VTRTVREALIMLEIKGLIVEVRGAGIYVLDSSGHADSDANVCNDAGFELLOAR
QLLESNIABFAALQATREDIVMKAQLEERELASSAPSGESGDMQFHLAIAEATH
NSMLVBLFRSQWRENNPMWIIQLSHLDLSDLYKRWLGDHKQILAAALIKKDAKAKL
AMQHLNENKQRLLEFNSNVDIYFDGYLFDSPWLDKVDK"
5953. .6345
/gene="yjid"
/locus_tag="c5405"
5953. .6345
/gene="yjid"
/locus_tag="c5405"
/note="Escherichia coli K-12 ortholog: b4326; Escherichia
coli O157:H7 ortholog: z5925"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yjid"
/protein_id="AAN83827.1"
/db_xref="GI:26111646"
/translation="MRKMQSVQTVLPESGNNTLSLRSDVCRDLFOLFSPHSPPLP
ILLVSGPWEQGNQSKLQSWYCRQLRSALLFHPRIAAALQVNLKRAYCHELAISL
EMMLYHDDLEPLTFDLVWQKSGSWHRTMPQ"
complement(6393. .6734)
/locus_tag="c5406"
complement(6393. .6734)
/locus_tag="c5406"
/note="Residues 1 to 113 of 113 are 95.57 pct identical to
residues 424 to 536 of 536 from Genpept.129 :
>gb|AA161895.1| (AF289032) putative
phosphoenolpyruvate-protein phosphoryltransferase Ppte
[Escherichia coli]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN83828.1"

gene

CDS

gene

CDS

/gene="uxuA"
/locus_tag="c5402"
1835. .3019
/gene="uxuA"
/locus_tag="c5402"
/EC_number="4.2.1.8"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="Escherichia coli K-12 ortholog: b4322; Escherichia
coli O157:H7 ortholog: z5920"
/codon_start=1
/transl_table=11
/product="Mannanase dehydratase"
/protein_id="AAN83824.1"
/db_xref="GI:26111643"
/translation="NEQTRWYGNPDSVLADVRQAGATGVVTAHLHI PNCEVSVSEI
LTKRAIVEDAGLVSWSVSEVPIHEDIKTHGVNEQWYIANYQOITRNLAQGITVICY
NFMVLDWTRTLDLEYVLPDGSKALRFQDIEFAAFEMHILKRPGEADTEEEIAQAE
RFATMSDSDKARLTENIYAGLPGAEGYTLQFRKHELYKDIIDAKLRNFAVFLKA
IIPVAEYVAGVHPPDPPIPLGPIRVSTIEDMOWVDTVNSMANGFTMCTGYSY
VRANDVDMIKQGPRIYFTHLRSTMEEDNPKTPHEAHLNGDMDVYEVKAIVEER
HRRKAEKEDLIPMRPDGHQMLDDKKKTNPGYSAIGRLKGLAEVRGVELAIQRAFF
SR"
3100. .4560

gene

CDS

gene

```
/db_xref="GI:26111647"
/translation="NIEVPAAMVIADKLASEVDPFISICTNDLTQYIMADKGNSTVAK
LVYCNDAVINAIAMVCQGRNEIIPVSCGEMAGDIQQTARLLTMGIDKLSASPSRL
PALKAARTSH"
complement(6771..7760)
/gene="yje"
/locus_tag="c5407"
complement(6771..7760)
/gene="yje"
/locus_tag="c5407"
complement(6771..7760)
/function="putative regulator"
/note="Escherichia coli K-12 ortholog: b4327; Escherichia
coli O157:H7 ortholog: z5926"
/codon_start=1
/transl_table=11
/product="Hypothetical transcriptional regulator yje"
/protein_id="AAN83829.1"
/db_xref="GI:26111648"
/translation="MKRKRLSAARMDEVCPOGRLETVNVMDDCGAVLHNIETKWLVD
FLTEKCRFSAQVSRVSPAFERRIRALEQALGVLEFNRQVTPLOLSQGGKLFHS
QIRHLQQLSRLASRGSDYQAKRIKIAAHSLGILFLSISQMPFLTWIAIEAI
DVEAVDKLRBQSDCIFFHDEDLLEAPFHIRLFESQLFVPCASDGHGALFNLAQ
PHFPLANTSRYSNGLINRLTRHSELSFSTFFVSSNSELKQVALDGGGIALMPEY
ATQQRISQQLVLRNDELVIPIQAYAYRMNTRMPVAERFWRLELEIPG"
complement(7747..8919)
/gene="iada"
/locus_tag="c5408"
complement(7747..8919)
/gene="iada"
/locus_tag="c5408"
/EC_number="3.4.19.-"
/function="enzyme; Degradation of proteins, peptides,
glyco"
/note="Escherichia coli K-12 ortholog: b4328; Escherichia
coli O157:H7 ortholog: z5927"
/codon_start=1
/transl_table=11
/product="Isoaspartyl dipeptidase"
/protein_id="AAN83830.1"
/db_xref="GI:26111649"
/translation="MIDYSAAGFTLLQGAHLYAPEDRGICDVLIAANGKIIIVASNIPS
DIVPCTVDLSGQILCPGFIQHVHLGGGEAGPTTRTEVALSRITAEAGITSVG
LIATDSISRHPESLLAKRALNEBISAWMLTGAHVPSRTITSGVSKDVAIIDRVIG
VKALSDRPSAPDQVYHLANMAESRVGLLGGKPGVTFHMGDSKALQVPVLDLEN
CDVPISKLPTHVNRNLFPEQALFARKGTIDITSSIDEFVAPAEIGIARVQAGIP
LARVLSDGNSQPFQDEGNLHIGVAGFETLLETQVQLVKDYFSLSQALRPLTS
SVAGFLNLTGKEILPGNDADLLVMTPELRISQVYARKLVMDKACVKGTFFETA"
Query Match 98.4%; Score 912.6; DB 1; Length 86898;
Best Local Similarity 99.0%; Pred. No. 1.6e-257;
Matches 918; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GTCATGTTAGCGCGATCATGTTAAATTTGCAGCATGCATCAGCAGGTCAGGACATTTGT 60
DB 85652 GTCATGTTAGCGCGATCATGTTAAATTTGCAGCATGCATCAGCAGGTCAGGACATTTGT 85593
QY 61 ACTTCTGTTTCGATTTAGTTCGCAATTTAGTCGCAACATGCAGACCCCGCATCTT 120
DB 85592 ACTTCTGTTTCGATTTAGTTCGCAATTTAGTCGCAACATGCAGACCCCGCATCTT 85533
QY 121 TATCGTTGAAGACGAGTTCGTTACACGCAACACGTTTGAAGAGTATTTTCGAAGCGGAGG 180
DB 85532 TATCGTTGAAGACGAGTTCGTTACACGCAACACGTTTGAAGAGTATTTTCGAAGCGGAGG 85473
QY 181 CTATGATGTTTTCGAAGCAGCATGTCGCGGAAATGCAATCAGATCTCTTCTGAATATGA 240
DB 85472 CTATGATGTTTTCGAAGCAGCATGTCGCGGAAATGCAATCAGATCTCTTCTGAATATGA 85413
QY 241 CATCAACCTGGTATCATGATATCAATCTCGCGGTAGAACGCTCTCTCTTAGCGCG 300
DB 85412 CATCAACCTGGTATCATGATATCAATCTCGCGGTAGAACGCTCTCTCTTAGCGCG 85353
QY 301 TGAACCTGCGGACGAGCGGAATGTTGCGTGTGATGTTCTCTGACTGCGCGGTGACAAACGAAGT 360
```

```
Db 85352 TGAACCTGCGGACGAGCGGAATGTTGCGTGTGATGTTCTCTGACTGCGCGGTGACAAACGAAGT 85293
QY 361 CGATAAAATTTCTCGGCTTCGAAATCGGTGAGATGACTATCATCAACCAACCGTTCAACCC 420
DB 85292 CGATAAAATTTCTCGGCTTCGAAATCGGTGAGATGACTATCATCAACCAACCGTTCAACCC 85233
QY 421 CGGTGAACTGACGATTCGTGACGCAACCTACTGTCGCGTACCATGATCTGGGTACTGT 480
DB 85232 CGGTGAACTGACTATTCGTGACGCAACCTGCTGTCGCGTACCATGATCTGGGTACTGT 85173
QY 481 CAGCGAAGAACGTCGTAGCGTTGAAAGCTCAAAAGTTCAATGTTGGGAACCTGGACATCAA 540
DB 85172 CAGCGAAGAACGTCGTAGCGTTGAAAGCTCAAAAGTTCAATGTTGGGAACCTGGACATCAA 85113
QY 541 CAGCGTTCGTGATTCGCGCCCTGATGCGGACAGTACAAAGTCGCGCGGAGGAGTTCCG 600
DB 85112 CAGCGTTCGTGATTCGCGCCCTGATGCGGACAGTACAAAGTCGCGCGGAGGAGTTCCG 85053
QY 601 CGCCATGCTTCACCTTCGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTCTGAA 660
DB 85052 CGCCATGCTTCACCTTCGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTCTGAA 84993
QY 661 GAAATGACCGCGCGTTCGAGCTGAAACCCAGCAGCTACTCTGTAGACGTGACGATCCGCG 720
DB 84992 GAAATGACCGCGCGTTCGAGCTGAAACCCAGCAGCTACTCTGTAGACGTGACGATCCGCG 84933
QY 721 TATTCGTAACATTTTCGATCTAGCCGGATACGCCGGAATCATCGCCACCATTCACGG 780
DB 84932 TATTCGTAACATTTTCGATCTAGCCGGATACGCCGGAATCATCGCCACCATTCACGG 84873
QY 781 TGAAGTTATCGCTTCGCGGTGATCTGGAAGATTAATCGGCTTTTACCACCGTCAAAAA 840
DB 84872 TGAAGTTATCGCTTCGCGGTGATCTGGAAGATTAATCGGCTTTTACCACCGTCAAAAA 84813
QY 841 AACGCGCTTTTACGCGCGTTTTTATTTTCAACCTTATTTTCCAGATACGTAACATCATC 900
DB 84812 AACGCGCTTTTACGCGCGTTTTTATTTTCAACCTTATTTTCCAGATACGTAACATCATC 84753
QY 901 GTCCGTTGTAACCTTCCTTTACTGGCTTT 927
DB 84752 GTCCGTTGTAACCTTCCTTTACTGGCTTT 84726
```

RESULT 7

```
LOCUS ECODYE 1468 bp DNA linear BCT 26-APR-1993
DEFINITION E.coli dye gene coding for Dye protein, complete cds.
ACCESSION M10044
VERSION M10044.1 GI:145817
KEYWORDS dye gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 1468)
AUTHORS Drury,L.S. and Buxton,R.S.
TITLE DNA sequence analysis of the dye gene of Escherichia coli reveals
amino acid homology between the dye and Ompr proteins
JOURNAL J. Biol. Chem. 260 (7), 4236-4242 (1985)
PUBMED 2984198
COMMENT
Original source text: E.coli K12 DNA, clone pB852.
Draft entry and sequence in computer readable, form kindly provided
by R.S.Buxton, 01-AUG-1985 [1].
Mutation of the dye gene results in sensitivity to dyes, envelope
protein changes, loss of expression of alkaline phosphatase, and
reduced transcription of sex factor F genes. The positions of the
dye and ompr genes on the K12 map are indicative of gene
duplication and the Dye protein is 28% homologous to the Ompr
protein.
A potential transcription termination site is located at positions
906-940 and a ribosome binding site at 161-166. An URF was found
on the complementary strand at positions >1542 to 948 (AA at 1541).
```

FEATURES	source	Location/Qualifiers
gene	1..1468	/organism="Escherichia coli"
		/mol_type="genomic DNA"
		/db_xref="taxon:562"
	98...814	
		/gene="dye"
CDS	98...814	
		/gene="dye"
		/codon_start=1
		/transl_table=11
		/protein_id="AAA23718.1"
		/db_xref="GI:145818"
		/translation="MQPFLHILVEDELVLTNLTLSIPKESGYDVFRATGDAEMHQILS EYDINGVIMDINLPGRKLLALRELEQANVLMFLTGRDNVDKILGLIGADDDYI KPNPRELITRARNLLSRNTNLGTVSEERSVSYKFNQWELDINSRLTGPQGEQVK LPSEFAMLHFENPKGKIOSRAELLKKMTKPHDRVTDVITIRRIKHPESTPD PELIATIHGSGYFCGDL"
ORIGIN	699 bp upstream of AccI site.	
	Query Match	98.4%; Score 912; DB 1; Length 1468;
	Best Local Similarity	99.9%; Pred. No. 1.1e-257;
	Matches 923; Conservative	0; Mismatches 0; Indels 1; Gaps 1;
Qy	5	TGTTACGCCGATCATGTTAAATTTCCAGCATGCATCAGCGAGGTGACGGGA-CTTTTGTACT 63
Db	1	TGTTACGCCGATCATGTTAAATTTGACATGCATCAGCGAGGTGACGGACCTTTTGTACT 60
Qy	64	TCCTGTTTTCGAATTTAGTTGGCAATTTAGGTAGCAACAATGCAGACCCCGCACATCTCTAT 123
Db	61	TCCTGTTTTCGAATTTAGTTGGCAATTTAGGTAGCAACAATGCAGACCCCGCACATCTCTAT 120
Qy	124	CGTTGAAGACGAGTGTGTTAACACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTA 183
Db	121	CGTTGAAGACGAGTGTGTTAACACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTA 180
Qy	184	TGATGTTTTTGAAGCGACAGATGCGCGGAAAATGCATCTCTCTGAAATATGACAT 243
Db	181	TGATGTTTTTGAAGCGACAGATGCGCGGAAAATGCATCAGATCTCTCTGAAATATGACAT 240
Qy	244	CAACCTCGTGATCATGATATCAATCTGCGGGTAAGAACGGTCTCTGTTAGCGGTGA 303
Db	241	CAACCTCGTGATCATGATATCAATCTGCGGGTAAGAACGGTCTCTGTTAGCGGTGA 300
Qy	304	ACTGCGCGAGCAGCGCGAATGTGCGTTGATGTTCTGACTGCGCGTGACAAACGAAGTCGA 363
Db	301	ACTGCGCGAGCAGCGCGAATGTGCGTTGATGTTCTGACTGCGCGTGACAAACGAAGTCGA 360
Qy	364	TAAATTTCTCGGCTCGAAATCGGTGCAGATGACTACATCAACAAACCGTTCAACCGCG 423
Db	361	TAAATTTCTCGGCTCGAAATCGGTGCAGATGACTACATCAACAAACCGTTCAACCGCG 420
Qy	424	TGAATTCACGATTCGTGCACGCAACCTACTGTCCCGTACCAATGAACTCTGGGTACTGTCAG 483
Db	421	TGAATTCACGATTCGTGCACGCAACCTACTGTCCCGTACCAATGAACTCTGGGTACTGTCAG 480
Qy	484	CGAAGAACGCTGTAGCGGTTGAAAGCTACAAGTTCAATGTTGGGAACTGCACATCAACAG 543
Db	481	CGAAGAACGCTGTAGCGTTGAAAGCTACAAGTTCAATGTTGGGAACTGCACATCAACAG 540
Qy	544	CCGTTCTGTGATCGGCCCTGATGCGGAGCAGTACAAGCTCGCGCGCAGCGAGTTCCGCGC 603
Db	541	CCGTTCTGTGATCGGCCCTGATGCGGAGCAGTACAAGCTCGCGCGCAGCGAGTTCCGCGC 600
Qy	604	CATGCTTCACTTCTGTGAAAACCCAGCGCAAAATTCAGTCCCGTCTGAACTGCTGAAGAA 663
Db	601	CATGCTTCACTTCTGTGAAAACCCAGCGCAAAATTCAGTCCCGTCTGAACTGCTGAAGAA 660
Qy	664	AATGACCGGCGTCAAGCTGAAACCGCACGACCGTACTGTAGAGTGTAAGCATCCGCGGTAT 723
Db	661	AATGACCGGCGTCAAGCTGAAACCGCACAGCCGTACTGTAGAGTGTAAGCATCCGCGGTAT 720
Qy	724	TCGTAAACATTTTCGAATTTACGCGGATACGCGGAAATCATCGCCACCACTTCAACGGTGA 783


```
to residues 60 to 241 of 241 from Escherichia coli K-12 :
B4311"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19417.1"
/db_xref="GI:30043698"
/translation="MESNTWNTIHDNKKENALNDVQVENVYAIKLDQWTVRPQGLT
HFSNGTRYGPVKLSWDATKDLNFGIRYRDWKAYRQODLSGDMRDNVHRWDGYVT
YHNSDFTFAWOTLLYSQNDYRYANHKWATENAFVLQYHMTPTDITPYEVDYLDRO
GVYNGRDLNLSNSYRIGVSFKL"
4029. .5243
/gene="yjht"
/locus_tag="S4469"
4029. .5243
/gene="yjht"
/locus_tag="S4469"
/notes="residues 1 to 404 of 404 are 98.76 pct identical to
residues 1 to 404 of 404 from Escherichia coli K-12 :
B4310"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19418.1"
/db_xref="GI:30043699"
/translation="MTTLTARVFTTAEIIRKTVIALVCHLNCRSQETVTWNKTIMAL
AIMMASFAANASVLPETVPFKSGTGADNDTVYIGLSAGTAWYKLDTOAKDKKWT
LAAPFGPREQATSAPIDGNLVFGGIGKNSGLTQVFNVDVHKYKPKNSVVKLMSHA
PMGAGHVTFFVHGKAVYTGVSQNIENGYFEDLNEAGKSDTAIDKINAHYFDKAE
YFFNFKLLSPDSTQNSYAGSPWGTAGAAVNVKGDITWMLNGEAKPGLRTDAVFE
DLTKSYNKLWNKLDVSPSGVAGGFAGISNDSLIFPAGGAGFKGSRNYQNGRYAHE
GLKKSYSTIHLHMHNGKWDKSGELSQGRAYGVSLPWNLSLIIGGETAGKRAVTDV
ISVKDKVTQVQ"
5530. .6288
/gene="yjh"
/locus_tag="S4470"
5530. .6288
/gene="yjh"
/locus_tag="S4470"
/notes="residues 1 to 252 of 252 are 99.60 pct identical to
residues 75 to 326 of 326 from Escherichia coli K-12 :
B4309"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19419.1"
/db_xref="GI:30043700"
/translation="MCGYHHPLATHQYQYGVQALHIARKLLPFIIPDNAGILIVPC
CRGSAFTAGSEGTYSERHGHASHDARWGTDTPLYODLVSRTAALAKNPQNKFLGVC
WMQGFEDLMTSDYASHPOHFHNMVEAFPRDLKQHSQNLNITDAPWFCGDTTWYK
PPHSYEATYGNQNNVLANIIFVDFQQOQERGLTNAPDEDDPDLSTGYGSAYRSP
WTTALRSSHFSFAARRGIISDRFVAILQFWRER"
6742. .7509
/notes="insertion element"
/insertion_seq="IS1X1"
/complement(6757. .7260)
/locus_tag="S4471"
/complement(6757. .7260)
/locus_tag="S4471"
/function="IS, phage, Tn; Transposon-related functions"
/notes="residues 1 to 167 of 167 are 89.82 pct identical to
residues 1 to 167 of 167 from GenPept :
>gb|AA044742.1|AF141323_13 (AF141323) Tnph (Shigella
flexneri)"
/codon_start=1
/transl_table=11
/product="IS1 orfB"
/protein_id="AAP19420.1"
/db_xref="GI:30043701"
/translation="MSROCTHYGRWPHQHGFTSLKKLRPQSVTSRIQPSDVIIVCAEMD
EHWGYGAKSRQRLFYADYDRIRTVVAHVGERLATLERLLLSLSPFVVMWTDG
WPLYESCLGKLKLVHSIKRYTQRIERNLNLKQHLARLVKRSLSFSKVELHDKVIGHY
LNKHYQ"
```

```
gene
CDS
complement(7179. .7454)
/locus_tag="S4472"
complement(7179. .7454)
/locus_tag="S4472"
/function="IS, phage, Tn; Transposon-related functions"
/notes="residues 1 to 91 of 91 are 98.90 pct identical to
residues 1 to 91 of 91 from GenPept :
>gb|AAK18550.1|AF348706_239 (AF348706) iso-IS1 orfA
[Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="IS1 orfA"
/protein_id="AAP19421.1"
/db_xref="GI:30043702"
/translation="WASISIRPCSAPEGVVRNGKSTAGHLRYLCSHCRKTKWQLQFT
YTASOPGTHQKIIDMANMGVGRASARIMGVGLNVLRLHKNSGRSR"
7549. .8244
/locus_tag="S4473"
7549. .8244
/locus_tag="S4473"
/function="regulatory"
/notes="residues 1 to 231 of 231 are 81.46 pct identical to
residues 30 to 261 of 261 from GenPept : >gb|AAI22652.1|
(AE008876) putative regulatory protein, deor family
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative deor family transcriptional regulator"
/protein_id="AAP19422.1"
/db_xref="GI:30043703"
/translation="MLEVSVMTIRRDLDHQBDEPLPLTLGGYIVMVKPAPSPMPVIHD
VPKRRDLPILAAAGVNNENDLIFFDNGQIEPLIVISMDAITFTGICYSHRVFA
LNEKPNVTAILCCGTYRARSDAFYDASNSPLDSLNPRIKIFISAGSVNHFVGFNFP
EDLATKRVNMRGLRKILLARHALFDEVASLASLAPISAFDVLISDRPLPADVYTHCQN
GSVKLIIPDSEDE"
complement(8345. .8461)
/locus_tag="S4474"
complement(8345. .8461)
/locus_tag="S4474"
/notes="residues 4 to 38 of 38 are 88.57 pct identical to
residues 29 to 63 of 96 from GenPept : >emb|CAC50649.1|
(AX188968) unnamed protein product [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19423.1"
/db_xref="GI:30043704"
/translation="MWAFARRALNVEDILCLPIQDCDKSHIWLWLVKDDQRLB"
Query Match 98.2%; Score 910.2; DB 1; Length 225944;
Best Local Similarity 99.6%; Pred. No. 1e-256;
Matches 923; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 GTCATGTTACGCCGATCATGTTAATTCGACATGATCAGGACAGGACATTTGT 60
|||||
Db 224698 GTCATGTTACGCCGATCATGTTAATTCGACATGATCAGGACAGGACATTTGT 224639
|||||
QY 61 ACTTCCTGTTTCGATTTAGTTGGCAATTTAGTAGCAACAATGACAGCCCGCACATTC 120
|||||
Db 224638 ACTTCCTGTTTCGATTTAGTTGGCAATTTAGTAGCAACAATGACAGCCCGCACATTC 224579
|||||
QY 121 TATCGTTTGAAGACGAGTTGGTAAACGCAACACCGTTCGAAAAGTATTTTCGAAGCGGAAG 180
|||||
Db 224578 TATCGTTTGAAGACGAGTTGGTAAACGCAACACCGTTCGAAAAGTATTTTCGAAGCGGAAG 224519
|||||
QY 181 CTATGATGTTTTGGAAGCGACAGATGGCGCGGAATATGATCATGATCCTCTCTGAATATGA 240
|||||
Db 224518 CTATGATGTTTTGGAAGCGACAGATGGCGCGGAATATGATCATGATCCTCTCTGAATATGA 224459
|||||
QY 241 CATCAACCTGGTGATCATGGATATCAATCTGCCGGGTGAAGACGGTCTCTCTGTAGCGCG 300
|||||
Db 224458 CATCAACCTGGTGATCATGGATATCAATCTGCCGGGTGAAGACGGTCTCTCTGTAGCGCG 224399
|||||
```


Qy	301	TGAACCTCGCGGAGCAGCGGAATGTTGCGTTGATGTTCTGACTGCGCGTGACCAAGAGT	360
Db	224338	TGAACCTCGCGGAGCAGCGGAATGTTGCGTTGATGTTCTGACTGCGCGTGACCAAGAGT	224339
Qy	361	CGATAAAATTTCTGCGCCTCGAAATCGGTGCGAGATGACTACATCAACCAACCGTTCAACCC	420
Db	224338	CGATAAAATTTCTGCGCCTCGAAATCGGTGCGAGATGACTACATCAACCAACCGTTCAACCC	224279
Qy	421	GGGTGAACAGCATGCTGCGACGCAACTACTGTCGCGTACCATGAAATCTGGGTACTGT	480
Db	224278	GGGTGAACAGCATGCTGCGACGCAACTGCTGTCGCGTACCATGAAATCTGGGTACTGT	224219
Qy	481	CAGCGAAGACCTGCTAGCGTTGAAAGCTTACAAGTTCAATGTTGCGAAGTGAACATCAA	540
Db	224218	CAGCGAAGACCTGCTAGCGTTGAAAGCTTACAAGTTCAATGTTGCGAAGTGAACATCAA	224159
Qy	541	CAGCGGTTGCTGTTGATCGCGCCCTGATGGCGGAGCAGTACAAGCTGCGCGCAGCGAGTTCCG	600
Db	224158	CAGCGGTTGCTGTTGATCGCGCCCTGATGGCGGAGCAGTACAAGCTGCGCGCAGCGAGTTCCG	224099
Qy	601	CGCCATGCTTCACTTCTGTGAAAACCCAGCGAAATTCAGTCCGCTGCTGAACCTGAA	660
Db	224098	CGCCATGCTTCACTTCTGTGAAAACCCAGCGAAATTCAGTCCGCTGCTGAACCTGAA	224039
Qy	661	GAATGACCGCGCGCTGAGCTGAAACCGCAGCAGCTAGTACGAGTGAACGATCCGCG	720
Db	224038	GAATGACCGCGCGCTGAGCTGAAACCGCAGCAGCTAGTACGAGTGAACGATCCGCG	223979
Qy	721	TATTCGTAAACATTTTCAATCTACGCGCGGATACGCGGAAATTCATCGCCACCATTCACGG	780
Db	223978	TATTCGTAAACATTTTCAATCTACGCGCGGATACGCGGAAATTCATCGCCACCATTCACGG	223919
Qy	781	TGAAGTTATCGCTTCTGCGGTGATCTGGAAGATTAAATCGGTTTACCGCGTCAAAAA	840
Db	223918	TGAAGTTATCGCTTCTGCGGTGATCTGGAAGATTAAATCGGTTTACCGCGTCAAAAA	223860
Qy	841	ACGCGCGCTTTTACGCGCGTTTATTTTCAACTTATTTCCAGATCAAGTCAATCATC	900
Db	223859	ACGCGCGCTTTTACGCGCGTTTATTTTCAACTTATTTCCAGATCAAGTCAATCATC	223800
Qy	901	GTCGGTGTGAATCTCTTACTCGCTTT 927	
Db	223799	GTCGGTGTGAATCTCTTACTCGCTTT 223773	
RESULT 10			
AL627284/c		88037 bp	DNA linear BCT 16-APR-2005
LOCUS			Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION			complete chromosome, segment 20/20.
ACCESSION			AL627284 AL513382
VERSION			AL627284.1 GI:16505618
KEYWORDS			Salmonella enterica subsp. enterica serovar Typhi
SOURCE			Salmonella enterica subsp. enterica serovar Typhi
ORGANISM			Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE			1 (bases 1 to 88037)
AUTHORS			Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.
TITLE			Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18
JOURNAL			Nature 413 (6858), 848-852 (2001)
PUBMED			11677608
REFERENCE			2 (bases 1 to 88037)
AUTHORS			Parkhill, J.
TITLE			Direct Submission

JOURNAL	Submitted (25-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT	E-mail: parkhill@sanger.ac.uk
FEATURES	Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
source	Location/Qualifiers 1. 88037 /organism="Salmonella enterica subsp. enterica serovar Typhi" /mol_type="genomic DNA" /strain="CT18" /db_xref="taxon:90370" 482. .1090 /gene="STY4858" 482. .1090 /gene="STY4858" /note="Similar to Streptomyces coelicolor hypothetical protein SC7H1.19 TR:O54178 (EMBL:AL021411) (182 aa) fasta scores: E(): 1.4e-10, 29.6% id in 169 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAD03347.1" /db_xref="GI:16505619" /db_xref="InterPro:IPR009412" /db_xref="UniProt/TREMBL:Q8Z0Y5" /translation="MKVITVTVPVQIRAKRCPSCSVKRDFTPSGAFRVSQKVLVD VMSYKTCYDCTWNISLFSRLPVSINRNLYCRIMDAATVQYFAYDNAILKRNNA ELSPDPDHIQERLVLSIASHKQSVSVRSFOVLSLSILKKQLLSAABIKERRIE TQQLSGVTMMLKSKRLKNKYDLQLSVETLYDRRRIIVLTRR" 1451. .2362 /gene="STY4859" 1451. .2362 /gene="STY4859" /note="N-terminus is similar to Salmonella typhimurium IICA i1CA TR:Q9RBP6 (EMBL:AF164435) (74 aa) fasta scores: E(): 3e-27, 98.6% id in 74 aa, indicated by promoterless reporter gene insertion to be induced upon bacterial invasion of phagocytic and nonphagocytic cells (but having no effect on virulence). The adjacent 70 aa is highly similar to the neighbouring CDS Salmonella typhimurium hypothetical protein TR:Q9RBP5 (EMBL:AF164435) (70 aa) fasta scores: E(): 1.9e-20, 91.4% id in 70 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAD03348.1" /db_xref="GI:16505620" /db_xref="UniProt/TREMBL:Q8Z0Y4" /translation="MHTPRYLRRHSDLYEDIPLEIREQILLIINTLGNCSFVDM TLYCVNHSDEVYRRICKTKRKYGLFTYAHSTSYLDEMSNLKKLDDKKRHIDTI ELAFNYIDTYLTVEVTLGLEPDKAISLNNIIFHEISLKYENGRIVLRRIKRLKN ICYLYSPGEYGFVEYDLMEAYNRLMDPACVRECHAVFRSLVIRHRSKIVVHE QDSLTNLMANMARGVISAQYVHKFHLIDVLESIFLPMAPEKSHHHAMLRISSE LACSYLTITERSIFLTQRAEDGVAP" 2573. .3034 /gene="STY4860" 2573. .3034 /gene="STY4860" /note="Similar to Neisseria meningitidis putative periplasmic protein NMA0993 TR:CAB84263 (EMBL:AL162754) (159 aa) fasta scores: E(): 1.2e-12, 34.4% id in 160 aa and Neisseria meningitidis hypothetical protein NNM0783 TR:AAF41196 (EMBL:AE002432) (159 aa) fasta scores: E(): 1.5e-12, 34.4% id in 160 aa. Also similar to Haemophilus ducreyi hypothetical protein TR:AAF33778 (EMBL:AP219260) (149 aa) fasta scores: E(): 9.2e-11, 31.5% id in 146 aa and Treponema pallidum hypothetical protein TP0895 TR:O83865 (EMBL:AE001259) (168 aa) fasta scores: E(): 6.9e-09, 33.6% id in 134 aa. Contains a probable

QY 601 CGCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTGCTGAA 660
DB 86191 CGCGATGCTGCACTTCTGCGAARACCTGGCAAAATTCAGTCTGCTGCGGAGCTGCTGAA 86132
QY 661 GAAATGACCGCGCTGAGCTGAAACCGCACGACCGTACTGTGACGTGACGATCCGCGC 720
DB 86131 AAAATGACCGGACGCGAGCTGAAGCGCATGACCGTACTGTGTGACGTCAACCATTCGTCG 86072
QY 721 TATTCGTAACATTTTCGATCTACCGCGATACCGCGGATACCGGAAATCATCGCCACCATTCACGG 780
DB 86071 TATTCGTAACATTTTCGATCTACCGCGATACCGCGGATACCGCGGAAATCATCGCCACCATTCACGG 86012
QY 781 TGAAGGTATTCGCTTCTGCGGTGATCTGGAAGATTAAT 818
DB 86011 CGAAGGTATTCGCTTCTGCGGTGATCTGGAAGATTAAT 85974

RESULT 11
LOCUS AE016849/c 284233 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 16
of 16 of the complete genome.
ACCESSION AE016849 AE014613
VERSION AE016849.1 GI:29140243
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi Ty2
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 284233)
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
12644504
REFERENCE 2 (bases 1 to 284233)
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 284233
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
301..489
/locus_tag="t4354"
301..489
/locus_tag="t4354"
/note="corresponds to STY4663 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO71807.1"
/db_xref="GI:29140244"
/translation="MTGKPSERHTGFIISGEMVVRDRCFNGEYLIHAGEAPVSEHNDA
WVGDTPCVALDFTHLR"
614..2125
/locus_tag="t4355"
614..2125
/locus_tag="t4355"
/locus_tag="t4355"
/note="corresponds to STY4664 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative DNA helicase"

gene

CDS

gene

CDS

/protein_id="AAO71808.1"
/db_xref="GI:29140245"
/translation="MPHTDTPQAAVIAWKGRLVVCFAFGSGKTTTLRRYAEENPTE
RMLYIANRAIRDEAEQFPNTCTKSHQLAWPTVCKGVSHRLINLSLFTVVARALN
SRNMLLARLLDLVLRNFCISAATDITFEHMPDDPCKGQPGQLFSAKKLWAMMSAR
QGDPEVHTDYLKLYQLSKNLSRSTYTLTLLPDEADANPVSAILVNLCTRWLVGDT
GQVIFRFGADNMQAPQLNSDRLLWLTSHFRFGPEVAGVANRLALKGETHQTITGK
GAGVLLTMPNFHGHAVLHRSVCGVIRTAHWSLAGRKVFWVGMSYRIEDLIDL
WFSIDMSERHRLTREYRDYDLQIAEDTGDVEMKQGI FILEQPPPLPDLRLTLR
EQRVISSEADVTTCYTAHRSKGLFMDRVKLDHDDLILDPMPSQRQDEINLMYAA
TRARKVLTIDPILHELLIQPRAEDIAGTISADKSSESVSTENDKKGTETVPES"
2109..3698
/locus_tag="t4356"
2109..3698
/locus_tag="t4356"
/note="corresponds to STY4665 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO71809.1"
/db_xref="GI:29140246"
/translation="MFRKVSWEKQKQPRQIEENKSNKAKRLIPTCTGYENPAISD
LLQHERQQWLRILWENSALPKEFRFYFLPLHGVGLCORLPASAKGKFAITDGMV
DYVITQTVFVRLSKGYLPRGASAEQSAQSVSAVVYAAALFHSGLRWLNIEGL
RSGAVWRPGLSVPEEPYRFRPFIAGVYGLIARLPELPVQLWQKQNPDLR
TLAPISGRYSAMDITDINAEIVHAGIPLGFPIAQEBOTSELABELNPAATVAPV
LRKSSSTPDPVPMVGQETVLLSSLDSDAKPERGEGQTAETDPLVQVMSILG
GLNTDQSPKEGQNTETNVAGESPEVALETSAAEIVINDVTTECTGSSSG
ISSLSSTELGDFWSLDLREGDIPVNTADACVHLTCGFVFI SVPGVFLFLKSHS
RSCSSGLKSGRKEQVQAAPKMKRHRVSDSRFPWQCCLYEPEGGRGKYLTYLIK
MSELYANGNPDDSLFLKVIN"
3862..4875
/locus_tag="t4357"
3862..4875
/locus_tag="t4357"
/note="corresponds to STY4666 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="probable phage integrase"
/protein_id="AAO71810.1"
/db_xref="GI:29140247"
/translation="METNITWQQLIDYFFAKPLRASEWSYTKVFKSFVHYMGPLSC
PNDTYHKVLAWRRFLLEKKSGLSGRTWNNKVHMRATFNQYIGQLLHYDENPNNVS
VKPDKKKTITQAIQVAYQIMSEYENQENTGLKYSCALPAPFMTVLTDLTY
TGIRQNQLHRLNDVLDREGQIRLITEGCKNKEHYVPVIFSRPLTCLMEKAQSE
GLKGNDRLFNIALFTGKDPALGDDMSQVRAFFRRLSKCQFALSPHRRFTLATSE
MKMPEONLHMAQSVLGHSHNMKSTLEYVENDIADVNGRALEQFMQIKAAHARSISGLT
KNR"
5302..5595
/locus_tag="t4358"
5302..5595
/locus_tag="t4358"
/note="corresponds to STY4667 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO71811.1"
/db_xref="GI:29140248"
/translation="MPVANSMAKMKRETINLRKIPARDLIDRAAKARGKNCITDFVLEA
ARAAERALISORIIIMADPEAVQAFILRLDQAPADNLRKTMQTQPAFWEQEK"
5592..6080
/locus_tag="t4359"
5592..6080
/locus_tag="t4359"
/note="corresponds to STY4668 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO71812.1"

Enterobacteriaceae: Salmonella.

REFERENCE
AUTHORS
1 (bases 1 to 13852)
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Porwollik, S., Ali, J., Pante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2
Nature 413 (6858), 852-856 (2001)
11677609
2 (bases 1 to 13852)
The Salmonella typhimurium Genome Sequencing Project
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
RegulonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES

Source

1. 13852
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

90..640
/gene="creA"
/note="synonym: STM4587"

-35_signal

90..98

/gene="creA"
/note="putative -35 signal for creA;
RegulonDB:STMLTH004682"

-10_signal

117..125

/gene="creA"
/note="putative -10 signal for creA;
RegulonDB:STMLTH004682"

156..161

/gene="creA"

/note="putative RBS for creA; RegulonDB:STMSIH004447"
167..640

/gene="creA"

/note="similar to E. coli orf, hypothetical protein
(AAC77350.1); Blastp hit to AAC77350.1 (157 aa), 89%
identity in aa 1 - 157"

/codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="AAL23402.1"

/db_xref="GI:16423154"

/translation="MKYKSLVLFSLILLGQARAEQIGSVDTVFVKMGFDPKHIVVEA
FDDPDVKNVTCYSRAKTGGKGLAEDTSDAISQOQVGPIELSDKIKNGKAQGE
VFKKRTSLIFKSLQVVRFDKRNKTLAYLAYSQVVDSPKNAISAVVPMWRQ"
641..1342

/gene="creB"

/note="synonym: STM4588"

641..646

/gene="creB"

/note="putative RBS for creB; RegulonDB:STMSIH004448"

653..1342

/gene="creB"

/note="OmpR family; similar to E. coli catabolic
regulation response regulator (AAC77351.1); Blastp hit to
AAC77351.1 (229 aa), 83% identity in aa 1 - 229"

/codon_start=1

/transl_table=11

/product="response regulator in two-component regulatory
system with CreC"

/protein_id="AAL23403.1"

/db_xref="GI:16423155"

/translation="MQQPQWLVEDEQGIADTLITLQEGTFVELFARGLPALAEKAR
QQRDAVLDVGLPDISGFCQLLERHPALPIFLTPARSDVDRLGLEIGADDYV
AKPFSREVSARVETLLRVRKFAAPSPPVTFHFLNEPAQAQIAWFTPLSLTRYEF
LLKTLKLSPEPVYSRQQLMDIVMSDAQETPDRTVDTHIKTLRAKRAINPELSPIT
HRGMGYSLRSV"

1330..2766

/gene="creC"

/note="synonym: STM4589"

1330..1335

/gene="creB"

/note="putative RBS for creC; RegulonDB:STMSIH004449"

1342..2766

/gene="creC"

/EC_number="2.7.3.-"

/note="senses catabolite repression; similar to E. coli
catabolite repression sensor kinase for PhoB; alternative
sensor for pho regulon (AAC77352.1); Blastp hit to
AAC77352.1 (474 aa), 80% identity in aa 1 - 474"

/codon_start=1

/transl_table=11

/product="sensory kinase (alternative) in two-component
regulatory system with CreB (or alternatively PhoB)"

/protein_id="AAL23404.1"

/db_xref="GI:16423156"

/translation="MRIGMRLLGLYFLVAVAVVLSIFVQEIKPGVRRATEGTLLD
TATLLAARDLDSGNPTDQLAKAFQLOHREFRANISIGIVKVRNHYHMYDAQG
KVLFDENKAVGDYSRWNDVQLTRGQYGARSTAKDAADPNSTVMYVAAPLISDGR
IGLVTVGKPAAMAPVTKSRRIIMASAVLLGALAIAGVMWMINRSIARLTIRYAD
SVTENRPLALGSLKRLQALESNRILEGKNIEQVYVALTHELKSPLAAIRG
AAETLRGGPADVVTPTENTLAQNTRMQALVETLLROALGNVLDNAIDFTPENGVITLS
TQLEARSQLAANKITLTLPSSIVVADAELLAQALGNVLDNAIDFTPENGVITLS
AQPMEGKAILQVDSGCCIPDFALPRIFDRFYSLPRENGRKSGLGLAFVSEARLLN
GEVTLNRPGEGLASLTLRHFT"

2812..4173

/gene="creB"

/note="synonym: STM4590"

2812..2817

/gene="creB"

/note="putative RBS for creB; RegulonDB:STMSIH004450"

2824..4173

/gene="creB"

/note="similar to E. coli tolerance to colicin E2

(AAC77353.1); Blastp hit to AAC77353.1 (450 aa), 73%

identity in aa 1 - 450"

/codon_start=1

/transl_table=11

/product="tolerance to colicin E2 protein"

/protein_id="AAL23405.1"

/db_xref="GI:16423157"

/translation="MLKSPLLWKMTLGGMILLILIPLMYVRHTIVRADYSHVEAA
IRQSTSGPKVGVPLVAIPVTELYTVLEEKEVKYKSLIFLFWLPESLVEGNQVVEA
RKIGYQGVWTDMAKAEFDVARLHNLNPNITLGPFFIVGVGDARGISAVKAPQ
VNGETLTVEPGTGLPEGREGIHIPLPDSQWATRLNLTLDMSLNLGFSFLVPGVRSS

AE017220_34	3400001	3510000	
AE017220_35	3500001	3610000	
AE017220_36	3600001	3710000	
AE017220_37	3700001	3810000	
AE017220_38	3800001	3910000	
AE017220_39	3900001	4010000	
AE017220_40	4000001	4110000	
AE017220_41	4100001	4210000	
AE017220_42	4200001	4310000	
AE017220_43	4300001	4410000	
AE017220_44	4400001	4510000	
AE017220_45	4500001	4610000	
AE017220_46	4600001	4710000	
AE017220_47	4700001	4755700	
Continuation (48 of 48) of AE017220 from base 4700001 (AE017220 Salmonella enterica subsp. enterica)			
Query Match 77.0%; Score 714; DB 1; Length 55700;			
Best Local Similarity 92.1%; Pred. No. 7.3e-199;			
Matches 753; Conservative 0; Mismatches 65; Indels 0; Gaps 0;			
QY	1	GTCAATGTTACGCGGATCATGTTAAATTTGCAGCATGCATCAGCAGGTTCAGGACTTTTGT	60
DB	54454	GTCAATGTTACGCGGATCATGTTAAATTTGCAGCATGCATCAGCAGGTTCAGGACTTTTGT	54395
QY	61	ACTTCTGTTTCGATTAGTTGGCAATTTAGGTAGCAACATGCAGACCCCGCACATTCT	120
DB	54394	ACTTCTGTTTCGATTAGTTGGCAATTTAGGTAGCAACATGCAGACCCCGCACATTCT	54335
QY	121	TATCGTTGAAGCAGATTGGTAACACGCAACATCGTTGAAAAGTATTTTGAAGCGGAAGG	180
DB	54334	TATCGTTGAAGCAGATTGGTAACACGCAACATCGTTGAAAAGTATTTTGAAGCGGAAGG	54275
QY	181	CTATGATGTTTTCGAAGCAGATGCGCGGAATGCATCAGATCTCTGTAATATGA	240
DB	54274	CTATGATGTTTTCGAAGCAGATGCGCGGAATGCATCAGATCTCTGTAATATGA	54215
QY	241	CATCAACCTGGTGATCATGATATCAATCTGCGGTAAGACGGTCTTCTGTAGCGG	300
DB	54214	CATCAACCTGGTGATCATGATATCAATCTGCGGTAAGACGGTCTTCTGTAGCGG	54155
QY	301	TGAACCTGCGAGCAGCGGAATGTTGCGTTGATGTTCTTGACTGCGCGTGACCAAGT	360
DB	54154	TGAACCTGCGAGCAGCGGAATGTTGCGTTGATGTTCTTGACTGCGCGTGACCAAGT	54095
QY	361	CGATAAATTTCTCGGCTCGAATCGGTGCAGATGACTACATCAACCAACCGTTCAACCC	420
DB	54094	CGATAAATTTCTCGGCTCGAATCGGTGCAGATGACTACATCAACCAACCGTTCAACCC	54035
QY	421	CGGTGAACCTGAGATTCGTGACGCAACCTACTGTCCTGATCATCATGATCTGGTACTGT	480
DB	54034	ACGCGAATTTGACATCCGTGCGCGCAACCTGTCCTGATCATCATGATCTGGTACTGT	53975
QY	481	CAGCGAAGACCTGCTAGCGTTGAAAGCTACAAGTTCAATGTTGGGAACCTGCACAA	540
DB	53974	CAGCGAAGACCTGCTAGCGTTGAAAGCTACAAGTTCAACGGTTGGGAACCTGCATCAA	53915
QY	541	CAGCGCTTCGTTGATCGGCGCTGATGCGGACGATGACAGTTCGCGCGCAGGATCCG	600
DB	53914	CAGTGTTCCTGATCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATG	53855
QY	601	CGCCATGCTTCACTTCTGTGAAACCCAGCGCAAAATTCAGTCCCGTGCATGCTGAA	660
DB	53854	CGCGATGCTTCACTTCTGTGAAACCCAGCGCAAAATTCAGTCCCGGCGAGCTGTGAA	53795
QY	661	GAAATGACCGCGCTGAGTGAACCCGACGACCTGCTAGACGATGACGATCCGCG	720
DB	53794	GAAATGACCGCGCGAGCTGAAGCGCATGACCGTACTGTGAGTTACCATTCGTG	53735
QY	721	TATTCGTAACATTTGATCTACCGCGATGACCGGGAATCATCGCCACCATTCACGG	780
DB	53734	TATTCGTAACATTTGATCTACCGCGATGACCGGGAATCATCGCCACCATTCACGG	53675
QY	781	TGAAGGTTATCGCTTCTCGGCTGATCTGGAAGATTAAAT	818

Db	53674	CGAAGTTATCGCTTCTGCGCGACCTGCAGGATTAAAT	53637
RESULT 14			
CP000026_45/c			
WPCOMMENT			
Sequence split into 46 fragments LOCUS CP000026 Accession CP000026			
	Fragment Name	Begin	End
	CP000026_00	1	110000
	CP000026_01	100001	110000
	CP000026_02	200001	310000
	CP000026_03	300001	410000
	CP000026_04	400001	510000
	CP000026_05	500001	610000
	CP000026_06	600001	710000
	CP000026_07	700001	810000
	CP000026_08	800001	910000
	CP000026_09	900001	1010000
	CP000026_10	1000001	1110000
	CP000026_11	1100001	1210000
	CP000026_12	1200001	1310000
	CP000026_13	1300001	1410000
	CP000026_14	1400001	1510000
	CP000026_15	1500001	1610000
	CP000026_16	1600001	1710000
	CP000026_17	1700001	1810000
	CP000026_18	1800001	1910000
	CP000026_19	1900001	2010000
	CP000026_20	2000001	2110000
	CP000026_21	2100001	2210000
	CP000026_22	2200001	2310000
	CP000026_23	2300001	2410000
	CP000026_24	2400001	2510000
	CP000026_25	2500001	2610000
	CP000026_26	2600001	2710000
	CP000026_27	2700001	2810000
	CP000026_28	2800001	2910000
	CP000026_29	2900001	3010000
	CP000026_30	3000001	3110000
	CP000026_31	3100001	3210000
	CP000026_32	3200001	3310000
	CP000026_33	3300001	3410000
	CP000026_34	3400001	3510000
	CP000026_35	3500001	3610000
	CP000026_36	3600001	3710000
	CP000026_37	3700001	3810000
	CP000026_38	3800001	3910000
	CP000026_39	3900001	4010000
	CP000026_40	4000001	4110000
	CP000026_41	4100001	4210000
	CP000026_42	4200001	4310000
	CP000026_43	4300001	4410000
	CP000026_44	4400001	4510000
	CP000026_45	4500001	4585229
Continuation (46 of 46) of CP000026 from base 4500001 (CP000026 Salmonella enterica subsp. enterica)			
Query Match 76.9%; Score 712.4; DB 1; Length 85229;			
Best Local Similarity 91.9%; Pred. No. 2.3e-198;			
Matches 752; Conservative 0; Mismatches 66; Indels 0; Gaps 0;			
QY	1	GTCAATGTTACCGCGATCATGTTAAATTTGCAGCATGCATCAGCAGGTTCAGGACTTTTGT	60
DB	83983	GTCAATGTTACCGCGATCATGTTAAATTTGCAGCATGCATCAGCAGGTTCAGGACTTTTGT	83924
QY	61	ACTTCTGTTTCGATTAGTTGGCAATTTAGGTAGCAACATGCAGACCCCGCACATTCT	120
DB	83923	ACTTCTGTTTCGATTAGTTGGCAATTTAGGTAGCAACATGCAGACCCCGCACATTCT	83864
QY	121	TATCGTTGAAGCAGATTGGTAACACGCAACATCGTTGAAAAGTATTTTGAAGCGGAAGG	180
DB	83863	TATCGTTGAAGCAGATTGGTAACACGCAACATCGTTGAAAAGTATTTTGAAGCGGAAGG	83804

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2006, 13:28:02 ; Search time 601 Seconds
(without alignments)

10279.817 Million cell updates/sec

Title: US-10-613-990A-31

Perfect score: 927

Sequence: 1 gtcgtgtaacgcgatcatg.....gtaacttttacttggttt 927

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*
- 14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	927	100.0	927	12	ADJ38402
c	925.8	99.9	1732	6	ABL90421 Escherich
3	577.2	62.3	726	11	ACH96017 Human pol
4	510.6	55.1	110000	10	ACF67367 28 Klebsiell
c	510.6	55.1	110000	10	ACF65386_0 Continuation (29 o
6	508.6	54.9	759	12	ADJ38390 Phototrab
7	493	53.2	717	10	ACF69844 Pantosa a
8	450.8	48.6	726	5	ADJ01748 Bacterial
c	447.4	48.3	1048	5	AAS92939 DNA encod
10	320.6	34.6	191996	13	ADT05647 Haemophil
c	319.4	34.5	110000	2	Continuation (10 o
12	316.2	34.1	711	4	AAS53409 Haemophil
13	316.2	34.1	711	8	ACA34287 Prokaryot
14	232.8	25.1	1110	5	AAS82422 DNA encod
15	161.6	17.4	2125	13	ADT05255 Haemophil
c	163.6	16.6	2227	13	ADT05194 Haemophil
17	137.4	14.8	729	6	ABQ90117 M. capsul
18	126	13.6	837	11	ABD17813 Pseudomon
c	126	13.6	2430	11	ABD17567 Pseudomon

c	20	126	13.6	3675	11	ABD17705
	21	120.6	13.0	990	11	ABD12913
	22	120.6	13.0	2166	11	ABD13118
	23	113.8	12.3	2004	11	ABD13261
	24	107.4	11.6	1018	3	AAA38554
c	25	106.8	11.5	11275	6	AAH43717
	26	103.8	11.2	714	8	ACA47574
	27	101.4	10.9	705	8	ACA48118
	28	100	10.8	100	8	ACD69947
	29	100	10.8	100	8	ACD69946
	30	100	10.8	100	8	ACD69948
	31	99	10.7	732	8	ACA28627
	32	98.6	10.6	110000	6	ABA03041_03
	33	98	10.6	714	8	ACA36700
	34	97.2	10.5	762	11	ABD07703
c	35	96.8	10.4	349980	6	ABQ81844
	36	96.2	10.4	705	6	ABN90763
	37	96.2	10.4	705	13	ADS03841
	38	96.2	10.4	721	8	ACA46767
	39	95	10.2	110000	6	ABQ69245_03
	40	95	10.2	110000	6	ABQ67197_02
	41	94.6	10.2	702	4	AAH52543
	42	94.6	10.2	3389	4	AAH54550
	43	93	10.0	699	4	AAS52039
	44	93	10.0	702	3	AAA95530
	45	93	10.0	702	13	ADZ74065

ALIGNMENTS

RESULT 1

ID	ADJ38402	standard; DNA; 927 BP.
XX		
AC	ADJ38402;	
XX		
DT	06-MAY-2004	(first entry)
XX		
DE	Escherichia coli arca (aerobic respiration control) gene SeqID31.	
XX		
KW	gamma-proteobacterium; Arca; aerobic respiration control;	
KW	Arca protein function; chromosomal arca gene; fermentation; L-amino acid;	
KW	L-lysine; L-glutamic acid; gene; ds.	
XX		
OS	Escherichia coli.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	101..817
FT		/*tag= a
FT		/product= "Escherichia coli arca protein"
XX		
PN	EPI382686-A1.	
XX		
PD	21-JAN-2004.	
XX		
PF	11-JUL-2003; 2003EP-00015911.	
XX		
PR	12-JUL-2002; 2002JP-00203764.	
XX		
PA	(AJIN) AJINOMOTO CO INC.	
XX		
PI	Ishikawa Y, Imaizumi A, Matsui K, Kojima H;	
XX		
DR	WPI; 2004-192975/19.	
DR	P-PSDB; ADJ38403.	
XX		
PT	New bacterium useful for producing a target substance (especially L-	
PT	lysine; L-glutamic acid or another L-amino acid) is modified so that the	
PT	Arca protein does not normally function.	
XX		
XX	Claim 6; SEQ ID NO 31; 38pp; English.	
XX		

CC This invention relates to a novel gamma-proteobacterium able to produce a
CC target substance such as an amino acid which is modified so that an Arca
CC (aerobic respiration control) protein does not normally function. The
CC disruption to Arca protein function is preferably due to disruption of a
CC chromosomal arca gene. The bacterium and method are useful in producing a
CC target substance by fermentation. The target substance is preferably an L
CC -amino acid, especially L-lysine or L-glutamic acid. Production of the
CC target substance using the modified bacterium is more efficient. The
CC present sequence is that of a PCR primer which was used for amplification
CC of the P ananatis arca gene in the exemplification of the invention.

XX Sequence 927 BP; 236 A; 228 C; 225 G; 238 T; 0 U; 0 Other;

Query Match	100.0%;	Score 927;	DB 12;	Length 927;
Best Local Similarity	100.0%;	Pred. No. 1.4e-297;		
Matches 927;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GTCTGTTTACGCGCATCATGTTAAATTTGCGATCATCATCAGGAGTTCAGGACTTTTGT	60	
Db	1	GTCTGTTTACGCGCATCATGTTAAATTTGCGATCATCATCAGGAGTTCAGGACTTTTGT	60	
Qy	61	ACTTCTGTTTGCATTTAGTTAGTGGCAATTAGGTAGCAAAACATGCGAGCCCGCACATTTCT	120	
Db	61	ACTTCTGTTTGCATTTAGTTAGTGGCAATTAGGTAGCAAAACATGCGAGCCCGCACATTTCT	120	
Qy	121	TATCTGTTGAGGAGGTTGGTAAACGACGACGACGTTGAAAGTATTTTCGAGCGGAGG	180	
Db	121	TATCTGTTGAGGAGGTTGGTAAACGACGACGACGTTGAAAGTATTTTCGAGCGGAGG	180	
Qy	181	CTATGATGTTTTCGAGGACGACAGATGGCGGCGGAAATGCGATCAGATCCTCTCTGAATATGA	240	
Db	181	CTATGATGTTTTCGAGGACGACAGATGGCGGCGGAAATGCGATCAGATCCTCTCTGAATATGA	240	
Qy	241	CATCAACTGTTGATCATGATATCAATCTGCGCGGTGAAGACGGTCTTCTGTTAGCGCG	300	
Db	241	CATCAACTGTTGATCATGATATCAATCTGCGCGGTGAAGACGGTCTTCTGTTAGCGCG	300	
Qy	301	TGAACCTGCGGAGGAGCGAATGTTGCGTTGATGTTCTGATGTCGTCGCGGTGACAGCAAGT	360	
Db	301	TGAACCTGCGGAGGAGCGAATGTTGCGTTGATGTTCTGATGTCGTCGCGGTGACAGCAAGT	360	
Qy	361	CGATAAAATTTCTCGGCTCGAAATCGGTGCGATGACTACATCAACCAACCGTTCAACCC	420	
Db	361	CGATAAAATTTCTCGGCTCGAAATCGGTGCGATGACTACATCAACCAACCGTTCAACCC	420	
Qy	421	GGGTGAACCTGACGATCTGTCGACGCAACTACTGTCGCGTACCATGAAATCTGGGTACTGT	480	
Db	421	GGGTGAACCTGACGATCTGTCGACGCAACTACTGTCGCGTACCATGAAATCTGGGTACTGT	480	
Qy	481	CAGCGAAGAACCTGCTAGGTTGAAAGCTTACAAAGTTCAATGTTGGGAACCTGGACATCA	540	
Db	481	CAGCGAAGAACCTGCTAGGTTGAAAGCTTACAAAGTTCAATGTTGGGAACCTGGACATCA	540	
Qy	541	CAGCGGTTCTGTTGATCGGCGCTGATGGCGAGCAGTACAAGTTCGCGCGCAGCGAGTTCG	600	
Db	541	CAGCGGTTCTGTTGATCGGCGCTGATGGCGAGCAGTACAAGTTCGCGCGCAGCGAGTTCG	600	
Qy	601	CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA	660	
Db	601	CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA	660	
Qy	661	GAATATGACCGCGCTGAGTCAAAACCGCAGCAGTCTAGTACGTTGAGACGATCCGCG	720	
Db	661	GAATATGACCGCGCTGAGTCAAAACCGCAGCAGTCTAGTACGTTGAGACGATCCGCG	720	
Qy	721	TATTCGTAACATTTTCGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG	780	
Db	721	TATTCGTAACATTTTCGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG	780	
Qy	781	TGAAGGTTATCGCTTCTCGGCTGATCTGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA	840	
Db	781	TGAAGGTTATCGCTTCTCGGCTGATCTGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA	840	

Qy	841	AACGGCGCTTTTACGCGCGCTTTTATTTTCAACCTTATTTCCAGATACGTAACATC	900
Db	841	AACGGCGCTTTTACGCGCGCTTTTATTTTCAACCTTATTTCCAGATACGTAACATC	900
Qy	901	GTCCGTTGTAACTTCTTTTACTGGCTTT	927
Db	901	GTCCGTTGTAACTTCTTTTACTGGCTTT	927

RESULT 2

ABL90421/c

ID ABL90421 standard; cDNA; 1732 BP.

XX ABL90421;

XX 24-MAY-2002 (first entry)

XX Human polynucleotide SEQ ID NO 983.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.

XX Homo sapiens.

XX WO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001WO-US016450.

XX 19-MAY-2000; 2000US-0205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Biree CE, Rosen CA;

XX WPI; 2002-122018/16.

XX P-PSDB; ABB90012.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders.

XX Claim 4; SEQ ID NO 983; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1732 BP; 439 A; 391 C; 417 G; 477 T; 0 U; 8 Other;

Query Match 99.9%; Score 925.8; DB 6; Length 1732;

Best Local Similarity 99.7%; Pred. No. 4.9e-297;

Matches 924; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTCTGTTTACGCGATCATGTTTAAATTTGCGATCATCAGCAGGTGAGGACTTTTGT 60
D 1313 GTCTGTTTACGCGATCATGTTTAAATTTGCGATCATCAGCAGGTGAGGACTTTTGT 1254
QY 61 ACTTCTGTTTTCGATTTAGTTGGCAATTTAGTAGCAACATGCGAGCCCGGCATTTCT 120
D 1253 ACTTCTGTTTTCGATTTAGTTGGCAATTTAGTAGCAACATGCGAGCCCGGCATTTCT 1194
QY 121 TATCGTTTGAAGACGAGTTGGTAACACGCAACACGTTTGAAGTATTTTGAAGCGGAAG 180
D 1193 TATCGTTTGAAGACGAGTTGGTAACACGCAACACGTTTGAAGTATTTTGAAGCGGAAG 1134
QY 181 CTATGATGTTTTCGAAGCAGACAGATGGCGCGGAATGCAATGCAATCTCTCTGAATATGA 240
D 1133 CTATGATGTTTTCGAAGCAGACAGATGGCGCGGAATGCAATGCAATCTCTCTGAATATGA 1074
QY 241 CATCAACTGGTGATCATGATATCAATCTGCGCGGTGAAGACGCTCTCTGTAGCGG 300
D 1073 CATCAACTGGTGATCATGATATCAATCTGCGCGGTGAAGACGCTCTCTGTAGCGG 1014
QY 301 TGAATCGCGCAGCAGCGCAATGTTGCGTTGATGTTCTCTGACTGGCGGTGACAAAGT 360
D 1013 TGAATCGCGCAGCAGCGCAATGTTGCGTTGATGTTCTCTGACTGGCGGTGACAAAGT 954
QY 361 CGATAAATTTCTCGGCTCGAAATCGGTGCGATGACTACATCAACAAACCGTTCAACCC 420
D 953 CGATAAATTTCTCGGCTCGAAATCGGTGCGATGACTACATCAACAAACCGTTCAACCC 894
QY 421 GGTGACAGTATGTCGACGCAACCTACTGTCCTGATGTTCTCTGACTGGCGGTGACAAAGT 480
D 893 GGTGACAGTATGTCGACGCAACCTACTGTCCTGATGTTCTCTGACTGGCGGTGACAAAGT 834
QY 481 CAGCGAAGACGTCGTAGCGTTGAAAGCTCAACAGTTCAATGTTGGGAATCGAATCAA 540
D 833 CAGCGAAGACGTCGTAGCGTTGAAAGCTCAACAGTTCAATGTTGGGAATCGAATCAA 774
QY 541 CAGCGCTTCGTTGATCGGCGCTTGATGCGGAGCAGTACAAAGCTGCGCGCAGCGAGTTCCG 600
D 773 CAGCGCTTCGTTGATCGGCGCTTGATGCGGAGCAGTACAAAGCTGCGCGCAGCGAGTTCCG 714
QY 601 CGCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTCTGNA 660
D 713 CGCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTCTGNA 654
QY 661 GAAATGACCGCGGTGAGCTGAAACCGCAGCAGTACTGTAGAGCTGACGATCGCGG 720
D 653 GAAATGACCGCGGTGAGCTGAAACCGCAGCAGTACTGTAGAGCTGACGATCGCGG 594
QY 721 TATTCGTAAACATTTTGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 780
D 593 TATTCGTAAACATTTTGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 534
QY 781 TGAAGTTATCGCTTCTCGGTGATCTGGAAGATTAAATCGGCTTTTACCAAGCTGAA 840
D 533 TGAAGTTATCGCTTCTCGGTGATCTGGAAGATTAAATCGGCTTTTACCAAGCTGAA 474
QY 841 AACGGCGCTTTTTCGCGCGTTTATTTTCAACCTATTTTCCAGATACGTAATCTATC 900
D 473 AACGGCGCTTTTTCGCGCGTTTATTTTCAACCTATTTTCCAGATACGTAATCTATC 414
QY 901 GTCCGTTGTAATCTTTTACTGCTTT 927
D 413 GTCCGTTGTAATCTTTTACTGCTTT 387
```

RESULT 3

ACH96017

ID ACH96017 standard; DNA; 726 BP.

XX ACH96017;

XX ACH96017;

XX 29-JUL-2004 (first entry)

XX

DE Klebsiella pneumoniae polynucleotide seqid 1812.

XX Recombinant expression vector; transcription regulatory element;

KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.

XX Klebsiella pneumoniae.

OS US610836-B1.

PN 26-AUG-2003.

PD 27-JAN-2000; 2000US-00489039.

PF 29-JAN-1999; 99US-0117747P.

PR (GENO-) GENOME THERAPEUTICS CORP.

PA Breton GL, Osborne M;

PI WPI; 2003-895346/82.

DR P-PSDB; ABO62466.

DR New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for preparing a vaccine composition against Klebsiella pneumoniae.

PT Disclosure; SEQ ID NO 1812; 932pp; English.

PS The invention describes a new isolated nucleic acid encoding a Klebsiella pneumoniae polypeptide. Also described are: a recombinant expression vector comprising the nucleic acid, operably linked to a transcription regulatory element; and a cell comprising the recombinant expression vector. The nucleic acid is useful for preparing a vaccine composition against Klebsiella pneumoniae. This sequence encodes a Klebsiella pneumoniae polypeptide of the invention

CC . Sequence 726 BP; 189 A; 198 C; 189 G; 150 T; 0 U; 0 Other;

Query Match 62.3%; Score 577.2; DB 11; Length 726;

Best Local Similarity 87.2%; Pred. No. 4.6e-181;

Matches 633; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 92 GTAGCAACATGACAGCCCGCAGATCTTATCGTTGAAGACGAGTTGGTAACAGCAAC 151

D 1 GTAGCAACATGACAGCCCGCAGATCTTATCGTTGAAGACGAGTTGGTAACAGCAAC 60

QY 152 ACGTTGAAAGTATTTTCGAAGCGAAGCTATGATGTTTCGAAGCGACAGATGCGCG 211

D 61 ACGTTGAAAGTATTTTCGAAGCGAAGCTATGATGTTTCGAAGCGACCGATGCGCG 120

QY 212 GAAATGATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGGATCAATCTG 271

D 121 GAAATGATCAGATCCTCTCTGAATATGATCAACCTGGTGATCATGGATCAACCTG 180

QY 272 CCGGTAAGAACGCTTCTGTAGCGGTGAACTGCGGAGCAGCGGATGTTGGTTG 331

D 181 CCGGTAAGAACGCTTCTGTAGCGGTGAACTGCGGAGCAGCGGATGTTGGTTG 240

QY 332 ATGTTCTTCTGACTGCGCGTGAACAAAGTTCGATAAAATTCCTGGCTCGAAATCGGTGA 391

D 241 ATGTTCTTCTGACTGCGCGTGAACAAAGTTCGATAAAATTCCTGGCTCGAAATCGGTGA 300

QY 392 GATGATCATATCACCACCAACCGTTCAACCGCGGTGAACTGACGATTCGTGACGCAACCTA 451

D 301 GATGATCATATCACCACCAACCGTTCAACCGCGGTGAACTGACGATTCGTGACGCAACCTG 360

QY 452 CTGTCCCGTACCATGATCTGGTACTGTACGCGAAGAACGTCGTAGCGTTGAAGACTAC 511

D 361 CTGTCCCGTACCATGATCTGGTACTGTACGCGAAGAACGTCGTAGCGTTGAAGACTAC 420

QY 512 AAGTTCAATGTTGGGAACCTGACATCAACCGGTTCGTTGATCGGCCCTCGATGCGGAG 571

D 421 AAGTTCAATGTTGGGAACCTGACATCAACCGGTTCGTTGATCGGCCCTCGATGCGGAG 480


```

XX Key Location/Qualifiers
FH CDS 41..757
FT /*tag= a "pantoea ananatis arcA protein"
FT /product=
XX
PN BP1382686-A1.
XX
XX 21-JAN-2004.
XX
XX 11-JUL-2003; 2003EP-00015911.
XX
XX 12-JUL-2002; 2002JP-00203764.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Ishikawa Y, Imaizumi A, Matsui K, Kojima H;
XX WPI; 2004-192975/19.
XX P-PSDB; ADJ38391.
XX
XX New bacterium useful for producing a target substance (especially L-
XX lysine, L-glutamic acid or another L-amino acid) is modified so that the
XX ArcA protein does not normally function.
XX
XX Example 4; SEQ ID NO 19; 38pp; English.
XX
XX This invention relates to a novel gamma-proteobacterium able to produce a
XX target substance such as an amino acid which is modified so that an ArcA
XX (aerobic respiration control) protein does not normally function. The
XX disruption to ArcA protein function is preferably due to disruption of a
XX chromosomal arcA gene. The bacterium and method are useful in producing a
XX target substance by fermentation. The target substance is preferably an L
XX -amino acid, especially L-lysine or L-glutamic acid. Production of the
XX target substance using the modified bacterium is more efficient. The
XX present sequence is that of the P ananatis gene which encodes the ArcA
XX protein and which was used in the exemplification of the invention.
XX
XX Sequence 759 BP; 205 A; 192 C; 183 G; 179 T; 0 U; 0 Other;
XX
XX Query Match 54.9%; Score 508.6; DB 12; Length 759;
XX Best Local Similarity 80.6%; Pred. No. 3.5e-158;
XX Matches 607; Conservative 0; Mismatches 144; Indels 2; Gaps 1;
XX
XX 69 TTTTCGATTAGTTGGC--AATTTAGGTAGCAACATGACAGCCCGCACATTTCTTTCGT 126
XX 7 TTTCAATTTAGTTGGCAAAATTTAGGTAGTAAATGACAGCCCGCACATTTCTTCATGT 66
XX
XX 127 TGAAGACGAGTTGGTAACACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTATGA 186
XX 67 TGAAGACGAACTGGTTCAGCGCAATACCTCAAAAGCATTTTGGAGCGGAAGGTTATGT 126
XX
XX 187 TGTTCGTAAGCAGACAGATGGCGCGGAATGATCAGATCCCTCTCTGTAATATGACATCAA 246
XX 127 CGTGACGAAGCGACCGATGGTGCAGAGATGCACAGGTGTTGACCGCAATGATGTCAA 186
XX
XX 247 CCGTGTGATCATGGATATCAATCTCGCGGTGAAGAACGCTCTTCTGTAGCGCGTAAGT 306
XX 187 TCTGGTTATTTAGGACATCAATCTCGCGGTGAAAACCGGCCCTGTTTACTGGCACGTAAGT 246
XX
XX 307 GCGGAGCAGCGGAATGTTGCGTTGATGTTCTCTGACTGGCCCGTGACAAAGAGTCGATAA 366
XX 247 GCGTGAGCAAGCAATGTGCGCATTTGATGTTCTCTGACCGGACGCGATACGGAATCGATAA 306
XX
XX 367 AATTCGCGCCTCGAATCGGTGCGAGATGACATCAACCAACCGTTCAACCCCGGTGA 426
XX 307 AATTCCTGGCGTGGAAATTTGGTGCAGACGACTACATTAAGCGCGCTTTAAACCCACGCA 366
XX
XX 427 ACTGACGATTCGTGCACGCAACCTACTGTCGCGTACCATGAAATCTGGGTACTGTGACGGA 486
XX 367 ATTAACATTCGTGCACGCTAACCTGCTGTGGCGACCAATGAATTTGCCCTTTACCAATGA 426
XX
XX 487 AGAAGCTCGTAGCGTTGAAAGCTACAAGTTCAATGTTGGGAACTGGGAACTCAACACAGCGG 546

```

RESULT 7

ACF69844

ID ACF69844 standard; DNA; 717 BP.

XX ACF69844;

AC ACF69844;

XX ACF69844;

DT 20-NOV-2003 (first entry)

XX Photorhabdus luminescens nucleotide sequence #8311.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 detection; food; gene expression; plant; animal; microorganism; toxin;
 antibiotic; biopesticide; virulence factor; disease model; plague;
 whooping cough; gene; ds.

XX Photorhabdus luminescens.

XX WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001PR-00001659.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

XX Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 Claim 2; SEQ ID NO 8311; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded

XX proteins from Photorhabdus luminescens. The isolated sequences are
 XX sources of probes and primers for detecting the genome of P. luminescens
 XX and related species; to study polymorphisms; for gene analysis and for
 XX detection/amplification of the genes. Antibodies (Ab) raised against the
 XX polypeptides encoded by the genes are used for detection/identification
 XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 XX carry a gene-containing vector are used to select compounds that
 XX modulate, regulate, induce or inhibit expression of the genes in plants,
 XX animals or microorganisms other than P. luminescens and are able to alter
 XX response or sensitivity to toxins and antibiotics produced by P.

CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX
XX Sequence 717 BP; 213 A; 158 C; 158 G; 188 T; 0 U; 0 Other;
Query Match 53.2%; Score 493; DB 10; Length 717;
Best Local Similarity 80.5%; Pred. No. 5.4e-153;
Matches 577; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 101 ATGCAGACCCGACATCTTATCGTTGAAGCAGAGTGGTAAACAGCAACGTTGAAA 160
Db 1 ATGCAACCCCGCACATTTTGATTGTTGAAGACGAAATGTCACTCGCAATACCCCTAATA 60
QY 161 AGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGACAGATGGCGCGAATGCAT 220
Db 61 AGCATTTTCGAAGCTGAAGGTTACATAGTTTATGAAGCACTGATGGTTGCAAAATGCAC 120
QY 221 CAGATCTCTCTGAAATATGACATCAACCTGTGTGATCATGGATPATCAATCTGCCGGTAA 280
Db 121 CATATCTGTCAACACAGACATTAACCTGTGTGATTATGATATTAACCTCCAGGCAAA 180
QY 281 AACGCTCTCTGTAGCGGTGAACCTGCGAGCAGGCGAATGTTGCGTTGATGTTCTGTG 340
Db 181 AATGCTTTGTACTTTGCCCGCAACTGCGAGAGCAAGCAAAATGTTGCTTTGATGTTCTGT 240
QY 341 ACTGCCGTGACACAGAGTGCATATAAATCTCGGCTCGAATCGTGCGAGATCACTAC 400
Db 241 ACTGCCGTGATACAGAGTGCATATAAATCTTGGCTTGGAAATCGTGCGAGATGATAC 300
QY 401 ATCAACAAACCGTTCAACCCGCGTGAACCTGACAGATTCGTGACGCAACCTACTGTCCGT 460
Db 301 ATCAACAAACCAATTAATCCGCGTGAATGACTATTCGCGCGCAACCTGCTTCTGT 360
QY 461 ACCATGATCTGGTACTGTGACGAGAAAGCTGTAGCGTTGAAAGCTACAAATGTTCAAT 520
Db 361 ACCATGATCTGAGTAATGTACGCGAGGAAACGTCGCCAAGTTGAGAGTTATAAATTCAC 420
QY 521 GGTTCGGAACCTGACATCAACAGCCGTTGTTGATCGGCCCTGTGATGGCGAGCAGTACAG 580
Db 421 GGTTCGGAATTAGATATCAATAGCCGTTCTCTGATAAGCCCGGCTGGTGAGCCGATATA 480
QY 581 CTGCGCGCAGCGAGTTCCGCGCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAG 640
Db 481 TTGCCACGCGAGGTTCCGCGCCATGCTTCACTTCTGTGAAATCCAGGCAAAATTCAG 540
QY 641 TCCCGTGTGAACTGCTGAGAAATAGACCGGCGCTGAGCTGAAACCGCAGACCGTACT 700
Db 541 ACACGCGCAGATTACTGAAAGAGATGACAGGTCGTGAAATGAAACCTCAGCAGCGTACT 600
QY 701 GTGACGCTGACGATCCGCGTATTTCGTAAACATTTTGAATCTACCGCGATACCGCGAA 760
Db 601 GTTGATGTAACGATTCGTCGTATCCGTAAACATTTTGAATCGACCCAGATACGATGAA 660
QY 761 ATCATCGCCACCATTCAGCGGTGAAGGTTATCGCTTCTCGGCTGATCTGGAAGATTA 817
Db 661 ATTATCGCCACCATTCAGCGGTGAAGGCTATCGTTCTGTGTTGATTTGAGAGCATAA 717

RESULT 8

ADFO1748

ID ADFO1748 standard; DNA; 726 BP.

XX

AC ADFO1748;

XX

DT 12-FEB-2004 (first entry)

XX Bacterial polynucleotide #2033.
DE
XX
XX
KW Proteus mirabilis infection; bacterial infection; antibacterial;
immunostimulant; gene; ds.
XX
XX
OS Proteus mirabilis.
XX
XX
PN US6605709-B1.
XX
XX
PD 12-AUG-2003.
XX
XX
PF 05-APR-2000; 2000US-00543681.
XX
XX
PR 09-APR-1999; 99US-0128706P.
XX
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX
PI Breton GL;
XX
XX
DR WPI; 2003-895291/82.
XX
XX
P-PSDB; ADF05920.
XX
XX
New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX
PS Disclosure; SEQ ID NO 2033; 870pp; English.
XX
XX
The invention relates to new Proteus mirabilis polypeptides and
polynucleotides. The invention also relates to antibodies against the
polynucleotides, methods for producing the polypeptides, a method of
generating vaccines for immunising an individual against P. mirabilis, a
method for evaluating a compound for the ability to bind a P. mirabilis
polypeptide and a method for screening test compounds for anti-bacterial
activity. The polypeptides and polynucleotides are useful as molecular
targets for diagnosing, preventing and treating pathological conditions
resulting from bacterial infection, as reagents for diagnosis of
bacterial diseases, as components of antibacterial vaccines, as targets
for antibacterial drugs or as bio-control agents for plants. This
sequence represents a Proteus mirabilis polynucleotide of the invention.
XX
SQ Sequence 726 BP; 224 A; 137 C; 149 G; 216 T; 0 U; 0 Other;

Query Match 48.6%; Score 450.8; DB 10; Length 726;

Best Local Similarity 76.3%; Pred. No. 6.3e-139;

Matches 554; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 92 GTAGCAACATGACAGACCCCGCACATCTTATCGTTGAAGACGAGTTGGTAAACAGCAAC 151
Db 1 GTAGCAATATGCAAAACCCCGCACATCTGATTTGGAAGATGAAGTAGTAGTCTGTAAT 60
QY 152 ACGTGTAAAAGTATTTTGAAGCGGAGGCTATGATGTTTTCGAAGCGACAGATGGCGG 211
Db 61 ACCCTGAAAAGCATATTGGAAGCTGAAGGGTATATCGTACACGAAGCCACTGATGGCAAC 120
QY 212 GAAATGATCAGATCTCTCTGATATGACATCAACCTGCTGATCATCATGATATCAATCTG 271
Db 121 GAGATGCAATAATATTCGTGCGACCAATGATATCAATCTGGTCATTATGATTAATCTT 180
QY 272 CCGGTTAAGAACGGTCTTCTGTTAGCGGCTGAACTGCGCGAGCAGCGGCAATGTTTCGGTTG 331
Db 181 CCTGGTAAAAATGGTCTTCTATTAGCCGCTGAATACGTGAACAGTGAAGTGTTCATTA 240
QY 332 ATGTTCTTGACTGGCGGTGACAAAGAGTGCATAAATTTCTCGGCTCGAATTCGGTGCA 391
Db 241 ATGTTCTTAACAGGTGCGTGAATAAGTGCATATAAATCTTAGGCTTTGAAATTTGGTGC 300
QY 392 GATGACTACATCAACCAACCGTTCAACCGGCTGAGACTGACGATTCGTCGACGCAACCTA 451
Db 301 GATGATTACATCACTAAACCACTTTAATCTCTGTTGAATTAACCTATCCGTGCTGTAACCTTA 360

QY 452 CTGCTCCGTACCATGAATCTGGTACTGTCTAGCGAAGAACGTCGTAGCGTTGAAAGCTAC 511
|||||
Db 361 TTGTCACGCACATGAATTTAGCGAATGGCAGAGAAGCGTCGTTTGTAGTTGAAGCTAT 420
512 AAGTTCAATGTTGGGAACTGGACATCAACAGCCGTTCTGTTGATCGGCCCTGTATGGCGAG 571
421 AAATTTAATGTTGGAGCTAGATATTAATAGTCGCTCTCTATTAGCCCTACAGGTGAA 480
572 CAGTACAAGCTGCCCGCAGCGAGTTCCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 631
481 CAGTATAAATACCTCGTAGTCGAGTTTCGTGCGATGTTACATTTCTGCGAAAACCCAGGA 540
632 AAAATTCACTCCGCTGCTGACTGCTGAAGAAATGACCGCGCTGACTGAAACCGCAC 691
541 AAAATCCAAACTCGTGCGAATTAATCTGAAAAAATGACGGGTCGTGAATTTAAACCTCAT 600
692 GACCTGACTGTAGAGCTGACGATCCGCGGTATTTCTGTAACATTTTCAATCTACGCGCGAT 751
601 GATCGTACTGTAGAGCTTACCATTCGTGCTATTCGTAAACATTTGAAATCAACCCCTGAT 660
752 AGCCCGGAAATCATCGCCACCAATTCACGGTGAAGGTTATCGCTTCTCGGCTGATCTGGAA 811
661 ACACCTGAGATTATCGCCACTATCCATGCTGAAGGTTATCGTTTCTGTGCTGATTTAGAC 720
QY 812 GATTAA 817
Db 721 GAGTGA 726

RESULT 9

AAS92939
ID AAS92939 standard; cDNA; 1048 BP.
XX
AC AAS92939;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28743.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG28752.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX
PS Claim 1; SEQ ID NO 28743; 103pp; English.

XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 1048 BP; 297 A; 265 C; 283 G; 203 T; 0 U; 0 Other;

Query Match 48.3%; Score 447.4; DB 5; Length 1048;
Best Local Similarity 99.8%; Pred. No. 1e-137;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 312 AGCAGGCGAATGTTGCGTTGATGTTCTCTGACTGCGCGTGACAAAGTGCATAAAATTC 371
Db 573 AACAGGCGAATGTTGCGTTGATGTTCTCTGACTGCGCGTGACAAAGTGCATAAAATTC 632
QY 372 TCGGCTCGAATCGGTGCGAGATGACTACATCACCACCAACCGTTCAACCCGGTGAACTGA 431
Db 633 TCGGCTCGAATCGGTGCGAGATGACTACATCACCACCAACCGTTCAACCCGGTGAACTGA 692
QY 432 CGATTGCTGCACGCAACCTACTGTCCTGACATGAAATCTGGTACTCTCAGCGAAGAAC 491
Db 693 CGATTGCTGCACGCAACCTACTGTCCTGACATGAAATCTGGTACTCTCAGCGAAGAAC 752
QY 492 GTCGTAGCGTTGAAAAGCTACAAAGTTCAATGTTGGGAACTCGACATCAACAGCCGTTCT 551
Db 753 GTCGTAGCGTTGAAAAGCTACAAAGTTCAATGTTGGGAACTCGACATCAACAGCCGTTCT 812
QY 552 TGATCGGCCCTGATGGCGAGCAGTACAAGTTCGCGCGCAGCGAGTTCCGCGCCATGCTTC 611
Db 813 TGATCGGCCCTGATGGCGAGCAGTACAAGTTCGCGCGCAGCGAGTTCCGCGCCATGCTTC 872
QY 612 ACTTCTGTGAAAACCCAGGCAAAATTCAGTCCCGTCTGAACTGCTGAAGAAAATGACCG 671
Db 873 ACTTCTGTGAAAACCCAGGCAAAATTCAGTCCCGTCTGAACTGCTGAAGAAAATGACCG 932
QY 672 GCCGTGAGCTGAAAACCCGACACCGGTACTGTAGAGCTGACGATCCGCGTATTCGTAAAC 731
Db 933 GCCGTGAGCTGAAAACCCGACACCGGTACTGTAGAGCTGACGATCCGCGTATTCGTAAAC 992
QY 732 ATTTGGAATCTACGCCGATACGCCGNA 760
Db 993 ATTTGGAATCTACGCCGATACGCCGNA 1021

RESULT 10

ADT05647/c
ID ADT05647 standard; DNA; 191996 BP.

XX
AC ADT05647;

XX
DT 02-DEC-2004 (first entry)

XX
DE Haemophilus influenzae (NTHi) DNA sequence - SEQ ID 693.

XX
KW middle ear bacterial infection; nasopharynx bacterial infection; ds.

XX
OS Haemophilus influenzae.

XX
PN WO2004078949-A2.

XX
PD 16-SEP-2004.

XX

PF 05-MAR-2004; 2004WO-US007001.

XX 06-MAR-2003; 2003US-0453134P.

XX (CHIL-) CHILDRENS HOSPITAL INC.

XX Bakaletz LO, Munson RS, Dyer DW;

XX WPI; 2004-662422/64.

XX New polynucleotides of nontypeable strain of Haemophilus influenzae,

XX useful for treating or preventing NTHi bacterial infections of the middle

XX ear and/or nasopharynx.

XX Claim 1; SEQ ID NO 683; 88bp; English.

XX The invention comprises nucleotide sequences (genes) from the genome of a

XX nontypeable strain of Haemophilus influenzae (NTHi). The NTHi DNA

XX sequences of the invention are useful for treating or preventing NTHi

XX bacterial infections of the middle ear and/or nasopharynx. The present

XX nucleic acid represents an NTHi DNA sequence of the invention.

XX Sequence 19196 BP; 59302 A; 35787 C; 37096 G; 59811 T; 0 U; 0 Other;

SQ Query Match 34.6%; Score 320.6; DB 13; Length 191996;

XX Best Local Similarity 66.0%; Pred. No. 3.7e-94;

XX Matches 480; Conservative 0; Mismatches 244; Indels 3; Gaps 1;

QY 85 AATTAGGTAGCAACATGACAGACCCCGACATTTCTTGAATATGACATCAACCTGGTGATCATGATAT 144

DB 55869 AGTATAGGATGGAATGACTACTCCAAATAATTCGTGTTGAAGATGAAATTTGTCTAC 55810

QY 145 AGCCAAACAGTTGAAAGATTTTCGAAGCGAAGCTATGATGTTTTCGAAGCGACAGA 204

DB 55809 TCGAATACGCTTAAGGATTTTGAAGCGAAGATGATGTTTGAAGCGAAGAA 55750

QY 205 TGGCGCGGAATGATCATGATCCTCTCTGAATATGACATCAACCTGGTGATCATGATAT 264

DB 55749 TGGTGTGAAATGCATATATATGCGCAATTCATAATTAATTTGTTGATGATAT 55690

QY 265 CAATCTGCGGTAGAACGCTTCTGTTAGCGGTGAACCTGCGGAGCAGCGCAATGT 324

DB 55689 TAAATTCAGGCAAAACCGCTTATTTATGCAAGAGAATCTCGTGAAGAAATTAAGCTT 55630

QY 325 TCGCTGTGATGTTCTGATGCGCGGTGACAAAGAGTCGATAAAATTCGCGCTCGAAT 384

DB 55629 ACCTCTTATTTTAACTGCTGCGAGATATGAGTGTGATAAAATTTAGTCTGGAAT 55570

QY 385 CGGTGAGATGACTACATCAACAAACCGTTCAACCCGCTGAACCTGACGATTCGTGCACG 444

DB 55569 TGGTCCGACGATTTTAAACCAACCTTTTAAACCTAGAGAATCTGATCCGCGCAG 55510

QY 445 CAACTACTGTCCTGATCATGAAATCTGGGTACTGTGCGAGAGAACGTCGTAGGCTTGA 504

DB 55509 TAACTATTTGATCGCACAATG---TCGCATCAAGAAAAGAAAATACATTTGTCGAGA 55453

QY 505 AAGCTACAAGTTCAATGTTTGGGAATCGACATCAACAGCCGTTCTGTTGATCGGCCCTGA 564

DB 55452 ATTCTATGTTTTAATGTTGGAATTTAGACCTAAATAGCCACAGTTTAAATACACAGA 55393

QY 565 TGGCGAGCAGTACAAGCTGCCGCGAGGAGTTCCGCGCCATGCTTCACTTCTGTGAAAA 624

DB 55392 AGGACAAGAAATTCAAACTTCTCGCAGTGAATTTCTGTCATGTTTACATTTCTGTGAAA 55333

QY 625 CCAGGCAAAATTTAGTCCGCTGTGAATCTGCTGAAGAAAATGACCGCGCGTGAAGTGA 684

DB 55332 TCCAGGAAAATTTGCAACCGCGAGAAATTTGCTGAAAAAATGACCGGACGAGGTTAAA 55273

QY 685 ACCGACGACGCTACTGTAGAGCTGACCATCCGCGTATTCGTAACATTTTCGATCTAC 744

DB 55272 ACCTCAGATGTCACCGTAGATGTACAAATTCGAGTATCAGAAACATTTTGAAGATCA 55213

QY 745 GCCGATACGCGGAAATCATCGCCACCATTTCAAGGATTTATCGTCTTCGCGTGA 804

DB 55212 TCCAAATACTCCAAATATCATTTATGACTATATCATGAGAGGCTATCGCTTTTGGGAGA 55153

QY 805 TCTGGAA 811

DB 55152 TATTGAA 55146

RESULT 11

AAT42063_09/c

Continuation (10 of 19) of AAT42063 from base 900001 (Haemophilus influenzae complete ge

WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

WP Fragment Name Begin End

WP AAT42063_00 1 110000

WP AAT42063_01 100001 210000

WP AAT42063_02 200001 310000

WP AAT42063_03 300001 410000

WP AAT42063_04 400001 510000

WP AAT42063_05 500001 610000

WP AAT42063_06 600001 710000

WP AAT42063_07 700001 810000

WP AAT42063_08 800001 910000

WP AAT42063_09 900001 1010000

WP AAT42063_10 1000001 1110000

WP AAT42063_11 1100001 1210000

WP AAT42063_12 1200001 1310000

WP AAT42063_13 1300001 1410000

WP AAT42063_14 1400001 1510000

WP AAT42063_15 1500001 1610000

WP AAT42063_16 1600001 1710000

WP AAT42063_17 1700001 1810000

WP AAT42063_18 1800001 1830121

Query Match 34.5%; Score 319.4; DB 2; Length 110000;

Best Local Similarity 65.8%; Pred. No. 6.8e-94;

Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTAGGTAGCAACATGACAGACCCCGACATTTCTTATCGTTGAAGACGAGTTGCTAAC 144

DB 36640 AGTATAGGATGGAATGACTACTCCAAAATTTCTCGTTGTTGAAGATGAATTTGTCTAC 36581

QY 145 ACGCAACACGTTGAAAAGTATTTTGAAGCGGAAGCTTATGATTTTTTGAAGCGACAGA 204

DB 36580 TCGAAATACGCTTAAAGGATTTTGAAGCGGAAGATATGATGTTTGAAGCAAGAAA 36521

QY 205 TGGCGGGAATGATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGATAT 264

DB 36520 TGGTGTGGAATGCAATCATATATTGGCAAAATCAATAATTAATTTGGTTGTGATGATAT 36461

QY 265 CAATCTGCGGTTAAGAACGGTCTTCTGTAGCGGTGAACTGCGCGAGCAGCGCAATGT 324

DB 36460 TAAATTTACAGGCAAAACGGCTTATTTATGCGCAAGAGATCTCGTGAAGATTAAGCTT 36401

QY 325 TGGCTGTGATGTTCTGATCGCGCGGTGACAAAGTCGATAAAATTTCTCGGCTCGAAAT 384

DB 36400 ACCTCTTATTTTAACTGCTGAGATAATGAAGTGGATAAATTTTAGGTTTGGAAAT 36341

QY 385 CGGTGAGATGACTACATCACCACCAACCGTTCAACCCGCGTGAACCTGACGATTCGTGCACG 444

DB 36340 TGGTGGCGATGACTATTTGACTAAGCCCTTTTAAACCTAGAGAATCTGACTATCCGCGCAG 36281

QY 445 CAACTACTGTCCTGATCATGAAATCTGGGTACTGTCTAGCGAAGAACCTGCTAGCGTTGA 504

DB 36280 TAACTATGATGCTGCAATG---CCGCAATCAAGAAAAGAAATACATTTGTTGTCGAGA 36224

QY 505 AAGCTACAAGTTCAATGTTGCGGAATCTGACATCAACAGCCGTTCTGTTGATCGGCCCTGA 564

DB 36223 ATTCTATCGTTTTAATGTTGGAATTTAGACCTTAAATAGCCACAGTTTAAATTTACACAGA 36164

QY 565 TGGCGAGCAGTACAAGCTGCGCGGAGGAGTTCCGCGCCATGCTTCACTTCTGTGAAAA 624

DB 36163 AGGACAAGAAATCAAACTTCTCGCAGTGAATTTCTGTCAGTGAATTTCTGTCGTAATTTCTGTGAAAA 36104

QY 625 CCCAGCAAAATTCAGTCCGCTGGAATCTGTGAAGAAATGACCGCCGCTGAGCTCAA 684
 |||||
 DB 36103 TCCAGAAATTCGAAACCGCGAAGATTCGTGAAAAAATGACCGACGAGGTTAAA 36044
 |||||
 QY 685 ACCGCACGACCGTACTGTAGAGCTGACGATCGCCGTTATTCGTAAACATTTGGAATCTAC 744
 |||||
 DB 36043 ACCTCAGGATCGTACCGTAGATGTCAAAATTCGAGGTATCAGAAACATTTTGAAGATCA 35984
 |||||
 QY 745 GCCGATAGCGCGAAATCATCGCCACCATTCACGCTGAAGTTATCGTCTCGGGTGA 804
 |||||
 DB 35983 TCCCAATATCTCAAAATATCATTTATGACTATATACATGGAAGGCTATCGTTTTTGTGGAGA 35924
 |||||
 QY 805 TCTGGAAGA 813
 |||||
 DB 35923 TATTGAATA 35915
 |||||

RESULT 12

AAS53409

ID AAS53409 standard; DNA; 711 BP.

XX

AC AAS53409;

XX

DT 13-FEB-2002 (first entry)

XX

DE Haemophilus influenzae DNA for cellular proliferation protein #191.

XX

KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
 antibacterial; drug design.

XX

OS Haemophilus influenzae.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US009180.

XX

PR 21-MAR-2000; 2000US-0191078P.

XX

PR 23-MAY-2000; 2000US-0206848P.

XX

PR 26-MAY-2000; 2000US-0207727P.

XX

PR 23-OCT-2000; 2000US-0242578P.

XX

PR 27-NOV-2000; 2000US-0253625P.

XX

PR 22-DEC-2000; 2000US-0257931P.

XX

PR 16-FEB-2001; 2001US-0269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX

PI Yamamoto RT, Xu HH;

XX

XX WPI; 2001-611495/70.

DR

DR P-PSDB; AAU35550.

XX

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX

PS Claim 27; SEQ ID NO 7046; 51lpp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes,
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence encodes an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences

XX Sequence 711 BP; 246 A; 113 C; 143 G; 209 T; 0 U; 0 Other;

Query Match 34.1%; Score 316.2; DB 4; Length 711;

Best Local Similarity 66.2%; Pred. No. 4.5e-94;

Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 101 ATGCAGACCCCGCACATTCCTTATCGTTGAAGACGAGTTGGTAACACGCAACACGTTGAAA 160

DB 1 ATGACTACTCCAAAAATTCCTCGTTTGAAGATGAATTTGTCACTCGAAATACGCTTAAA 60

QY 161 AGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGACAGATGGCGCGGAATGCAT 220

DB 61 GGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCAGAAAAATGGTGTGTTGAAATGCAT 120

QY 221 CAGATCCTCTCTGAATATGACATCAACCTGCTGATCATGGATATCAATCTGCCGGCTAAG 280

DB 121 CATATTTGGCAATCATTAATTAATTTGTTGATGGATATTAATTTACAGGCAAA 180

QY 281 AACGGTCTTCTGTAGCGCGTGAACCTGCGAGCAGGCGAAATGTTGCGTTGATGTTCCCTG 340

DB 181 AACGGCTTATTATTGGCAAGAGAACTCCGTGAAGAAATTAAGCTTACCTCTTATTTTATA 240

QY 341 ACTGCCGTGACAAAGAGTGCATAAATTTCTGGCCTCGAAATCGGTGCAGATCACTAC 400

DB 241 ACTGTCGAGATAAATGAAGTGGATAAATTTTAGGTTTGGAAATTTGGTCCGATGACTAT 300

QY 401 ATCAACCAACCGTTCAACCCGCGTGAACCTGACGATTCGTGACGCAACCTACTGTCCTG 460

DB 301 TTGACTAAGCCTTTTAACCCCTAGAGAACTGACTATCCGCGCAGGTAAATCTATTGTCAT 360

QY 461 ACCATGAATCTGGGTACTGTCTCAGCGAAGAACGTCGTAGCTGTGAAAGCTACAAGTTCAAT 520

DB 361 GCAATG---CCGCATCAAGAAAAAAGAAATACATTTTGGTCGAGAAATCTATCGTTTAA 417

QY 521 GGTGGGAATCTGGACATCAACAGCGGTTCTGTTGATCGGCCCTGATGGCGAGCAGTACAG 580

DB 418 GGTGGGAATTTAGACCTAAATAGCCACAGTTTAAATTACACCAGAAAGGACAAAGAAATCAA 477

QY 581 CTGCGCGCAGCGAGTTCCGCGCCATGCTTCTGTTGAAACCCAGGCAAAATTCAG 640

DB 478 CTTCCTCGCAGTGAATTTCTGTCGAATGTTACATTTCTGTGAAATTCAGGAAATTCGAA 537

QY 641 TCCCGTCTGAACTGCTGAAGAAATGACCGCGCTGAGCTGAAACCCGACGACCGTACT 700

DB 538 ACGCGGAAGAAATTCCTGAAAAAATGACCGGACGAGAGTTAAAAACCTCAGGATCGTACC 597

QY 701 GTAGACGTGACGATCCGCCGTTATTCGTAAACATTTGCAATCTACGCCGATACGCGGAA 760

DB 598 GTAGATGTCACAATTCGACGTTATCAGAAAAACATTTTGAAGATCATCCCAATACTCCAAAT 657

QY 761 ATCATCGCCACCATTCACGGTGAAGTTATCGCTTCTCGGGTGATCTGGAAGA 813

DB 658 ATCATTATGACTATACATGGAGAGGCTATCGTTTTTGTGGAGATTAATGAATA 710

RESULT 13

ACA34287

ID ACA34287 standard; DNA; 711 BP.

XX

AC ACA34287;

XX

DT 19-JUN-2003 (first entry)

XX

DE Prokaryotic essential gene #15944.

XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 18226; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1110 BP; 256 A; 291 C; 296 G; 267 T; 0 U; 0 Other;
Query Match 25.1%; Score 232.8; DB 5; Length 1110;
Best Local Similarity 98.8%; Pred. No. 3.6e-66;
Matches 245; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 544 CGTTTCGTTGATCGGCCCTGTATGGCGAGCAGTA-CAAGCTGCGCGCAGCGAGTTCCGCG 602
DB 761 CGTTTCGTTGATCGGCCCTGTATGGCGAGCAGTA-CAAGCTGCGCGCAGCGAGTTCCGCG 820
QY 603 CCATGCTTCACCTCTGTGTAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAAGA 662
DB 821 CCATGCTTCACCTCTGTGTAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAAGA 880
QY 663 AATGACCGCGCGTGAAGTGAACCGCAGCGCTGCTAGACCTGACGATCCGCGCTA 722
DB 881 AATGACCGCGCGTGAAGTGAACCGCAGCGCTGCTAGACCTGACGATCCGCGCTA 940
QY 723 TTCGTAACATTTGGAATCTACCGCGGATACCGCGGAATCATCGCCACCATTTCACGGTG 782
DB 941 TTCGTAACATTTGGAATCTACCGCGGATACCGCGGAATCATCGCCACCATTTCACGGTG 1000
QY 783 AAGGTTAT 790
DB 1001 AAGTTTAT 1008
RESULT 15
ADT05255
ID ADT05255 standard; DNA; 2125 BP.
XX AC
XX ADT05255;
XX
XX 02-DEC-2004 (first entry)
XX
XX Haemophilus influenzae (NTHI) contig DNA sequence - SEQ ID 291.
DE
XX middle ear bacterial infection; nasopharynx bacterial infection; ds;
KW contig.
XX
XX Haemophilus influenzae.
OS
XX WO2004078949-A2.
XX

PD 16-SEP-2004.
XX
XX 05-MAR-2004; 2004WO-US007001.
XX
XX 06-MAR-2003; 2003US-0453134P.
XX
XX (CHIL-) CHILDRENS HOSPITAL INC.
XX
XX Bakaletz LO, Munson RS, Dyer DW;
XX WPI; 2004-662422/64.
XX
XX New polynucleotides of nontypeable strain of Haemophilus influenzae,
PT useful for treating or preventing NTHi bacterial infections of the middle
PT ear and/or nasopharynx.
XX
XX Example 1; SEQ ID NO 291; 88pp; English.
XX
XX The invention comprises nucleotide sequences (genes) from the genome of a
CC nontypeable strain of Haemophilus influenzae (NTHi). The NTHi DNA
CC sequences of the invention are useful for treating or preventing NTHi
CC bacterial infections of the middle ear and/or nasopharynx. The present
CC nucleic acid represents an NTHi contig sequence of the invention.
XX
SQ Sequence 2125 BP; 596 A; 401 C; 462 G; 666 T; 0 U; 0 Other;
Query Match 17.4%; Score 161.6; DB 13; Length 2125;
Best Local Similarity 65.5%; Pred. No. 2.8e-42;
Matches 253; Conservative 0; Mismatches 129; Indels 4; Gaps 1;
QY 426 AACTGACGATTTCGTGCAGCAACCTACTGTCCCGTACCATGAATCTGGGTACTGTGAGG 485
DB 1 AACTGACGATTTCGTGCAGCAACCTACTGTCCCGTACCATGAATCTGGGTACTGTGAGG 56
QY 486 AAGAACGCTGTAGCGTTGAAAGCTCAAGTTCAAGTTGGTGGAACTGGACATCAACAGCC 545
DB 57 AAAATACATTTGGTTCGAGAAATTTCTATCGTTTAAATGGTTGGAATTAGACCTAAATAGCC 116
QY 546 GTTCGTTGATCGGCCCTGTATGGCGAGCAGTACAGCTGCGCGGAGCGAGTTCCGCGCA 605
DB 117 ACAGTTTAAATTACACCAAGAGGACAGAAATTCAAATCTCCTCGCAGTGAATTTCTGTCAA 176
QY 606 TGCTTCACCTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTCTGAACTGCTGAAGAAAA 665
DB 177 TGTACATTTCTGTGAAATTCAGGAAATTCGAAACCGCGGAGAAATTCGTGAAAAA 236
QY 666 TGAACCGCGCTGTAGCTGAAACCGGACGACCGTA CTGTAGAGCTGACGATCCGCGCTATTC 725
DB 237 TGAACCGGACGAGAGTTAAACCTCAGGATCGTACCGTAGATGTCAACAATTCGACGTATCA 296
QY 726 GTAAACATTTGGAAGATCATCCAAATCTCAAATATCATTTATGACTATACATGAGAGAG 785
DB 297 GAAACATTTTGAAGATCATCCAAATCTCAAATATCATTTATGACTATACATGAGAGAG 356
QY 786 GTTATCGCTTCTGCGGTGATCTGGAA 811
DB 357 GCTATCGCTTCTGCGGTGATCTGGAA 382
Search completed: February 13, 2006, 13:42:18
Job time : 606 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2006, 13:37:39 ; Search time 4551 Seconds
(without alignments)
9530.131 Million cell updates/sec

Title: US-10-613-990A-31

Perfect score: 927

Sequence: 1 gtcattgtacgcgatcatg.....gtaactttcttactggcttt 927

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gse1:*
10: gb_gse2:*
11: gb_gse3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	481.2	51.9	772	10	CL681756 PRI0132a
C 2	343	37.0	800	10	CL688716
C 3	114	12.3	114	2	BG834429 352648 MA
C 4	112	12.1	841	2	BG367933 HVSME1001
C 5	102	11.0	720	10	CNS00WTA
C 6	99.6	10.7	768	10	CL685986 PRI0142c
C 7	93.2	10.1	1193	9	BZ574214 msh2 3562
C 8	92.6	10.0	831	10	CL652351 PRI0114c
C 9	92.6	10.0	1018	9	BZ574107 msh2 3514
C 10	89.8	9.7	1159	9	BZ559286 pacs2-164
C 11	89.6	9.7	1575	9	BZ574473 msh2 3684
C 12	88.6	9.6	1029	9	BZ551846 pacs1-60
C 13	84.8	9.1	732	10	CL944656 O81PS8005
C 14	82.8	8.9	665	10	CW158500 104_584_1
C 15	81.4	8.8	1189	9	BZ564167 pacs2-164
C 16	81.4	8.8	1356	9	BZ579121 msh2 6208
C 17	78.6	8.5	733	8	CV982963 UMC-bmix
C 18	78.4	8.5	875	8	DR888097 JGI_CUNH1
C 19	78.4	8.5	956	8	CV983114 UMC-bmix
C 20	77.2	8.3	1347	9	BZ570026 msh2_1131
C 21	75.8	8.2	1166	9	BZ549106 pacs1-60
C 22	75.6	8.2	816	6	CB991419 AGENCOURT

C 23	75.6	8.2	865	6	CB994454
C 24	75	8.1	874	7	CN822966
C 25	74.8	8.1	312	2	BE323316
C 26	74.4	8.0	544	9	BH017051
C 27	74.2	8.0	807	6	CB998053
C 28	74.2	8.0	924	9	BZ549107
C 29	73.6	7.9	911	8	DN640699
C 30	73.4	7.9	628	9	BZ551497
C 31	72.6	7.8	667	2	B1263997
C 32	71	7.7	703	8	DN815843
C 33	71	7.7	768	7	CN821815
C 34	71	7.7	871	7	CN824699
C 35	70.2	7.6	847	9	BZ553164
C 36	70	7.6	568	5	BV696722
C 37	69.4	7.5	555	7	CV122221
C 38	69.4	7.5	594	7	CJ368177
C 39	69.4	7.5	697	7	CJ376658
C 40	68.6	7.4	1263	9	BZ553503
C 41	66	7.1	366	7	CO990188
C 42	66	7.1	748	7	CN205135
C 43	65.4	7.1	1274	9	BZ572156
C 44	65.2	7.0	848	10	CL688368
C 45	64.8	7.0	867	10	CZ531116

ALIGNMENTS

RESULT 1
CL681756/c

LOCUS

DEFINITION

CL681756 772 bp DNA linear GSS 09-JUL-2004
PRI0132a All 2 - PRI0132a.BR (772) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

CL681756

CL681756.1

GI:50188810

GSS.

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

1 (bases 1 to 772)

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.

AppADB: an AcedB database for the nematode satellite organism

Pristionchus pacificus

Nucleic Acids Res. 32 (1), D421-D422 (2004)

14681447

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end

sequenced at Vancouver, Canada.

Seg primer: T7

Class: fosmid ends.

Location/Qualifiers

1. .772

/organism="Pristionchus pacificus"

/mol_type="genomic DNA"

/strain="California"

/db_xref="taxon:54126"

/clone_lib="Mixed stage fosmid library of P. pacificus

var. California"

/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match 51.9%; Score 481.2; DB 10; Length 772;

Best Local Similarity 98.4%; Pred. No. 1.7e-140;

Matches 486; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
QY 434 ATTCGTGACGCAACTACTCTCCGTACCATGAATCTGGGTACTGTCTCAGCGAAGAACGT 493
DB 772 ATTCGTGACGCAACTCTGCTCCGTACCATGAATCTGGGTACTGTCTCAGCGAAGAACGT 713
QY 494 CTTAGCGTTGAAAGTACAAAGTTCAATGGTTGGGAACCTGGACATCAACAGCGCTTCGTTG 553
DB 712 CTTAGCGTTGAAAGTACAAAGTTCAATGGTTGGGAACCTGGACATCAACAGCGCTTCGTTG 653
QY 554 ATCCGCCCTGATGGCGAGCAGTACAAAGCTGCCGCGCAGCGAGTTCGCGCCATGCTTCAC 613
DB 652 ATCCGCCCTGATGGCGAGCAGTACAAAGCTGCCGCGCAGCGAGTTCGCGCCATGCTTCAC 593
QY 614 TTCTGTGAAGAACCCAGCGAAATTCAGTCCCGTGTCTGAACTGCTCAAGAAATGACCGGC 673
DB 592 TTCTGTGAAGAACCCAGCGAAATTCAGTCTCGTGTCTGAACTGCTCAAGAAATGACCGGC 533
QY 674 CTTAGCTGAAGAACCCGACGACCGTACTGTAGACGTGACGATCCGCGCTATTGCTAAACAT 733
DB 532 CTTAGCTGAAGAACCCGACGACCGTACTGTAGACGTGACGATCCGCGCTATTGCTAAACAT 473
QY 734 TTGGAATCTACGCCGATACGCCGGAATCATGCCACCATTCACGCTGAAGTTATCGC 793
DB 472 TTGGAATCTACGCCGATACGCCGGAATCATGCCACCATTCACGCTGAAGTTATCGC 413
QY 794 TTCTCGGCTGATCTGGAAGATTAATCGGCTTTTACCACCGTCAAAAAAAGCGCGCTTTT 853
DB 412 TTCTCGGCTGATCTGGAAGATTAATCGGCTTTTACCACCGTCAAAAAAAGCGCGCTTTT 353
QY 854 AGCGCGCTTTTATTTTCAACCTTTATTTCCAGATACGTAACCTCATCGTTCGTTGTAAC 913
DB 352 AGCGCGCTTTTATTTTCAACCTTTATTTCCAGATACGTAACCTCATCGTTCGTTGTAAC 293
QY 914 TCTTTACTGGCTTT 927
DB 292 TCTTTACTGGCTTT 279
```

```
RESULT 2
LOCUS CL686716
DEFINITION CL686716 800 bp DNA linear GSS 09-JUL-2004
PRIO144d_H03_2 - PRIO144d_BR (800) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL686716
VERSION CL686716.1 GI:50195255
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
REFERENCE Srinivasan,J., Otto,G.W., Kahlow,U., Gelsler,R. and Sommer,R.J.
AUTHORS AppADB: an AcedB database for the nematode satellite organism
TITLE Pristionchus pacificus
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
PUBMED 1468147
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spenannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
FEATURES
Location/Qualifiers
1..800
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
```

```
ORIGIN
Query Match 37.0%; Score 343; DB 10; Length 800;
Best Local Similarity 100.0%; Pred. No. 1e-96;
Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCATGTTACGCCCATCATGTTAATTTTCAGCATGTCATCAGGCAGGTTCAGGACTTTTGT 60
DB 458 GTCATGTTACGCCCATCATGTTAATTTTCAGCATGTCATCAGGCAGGTTCAGGACTTTTGT 517
QY 61 ACTTCCTGTTTCGATTTTAGTTGGCAATTTAGTAGCAAAATGATGACAGCCCGCACATCT 120
DB 518 ACTTCCTGTTTCGATTTTAGTTGGCAATTTAGTAGCAAAATGATGACAGCCCGCACATCT 577
QY 121 TATCGTTGAAGACGAGTTGGTAACAACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAG 180
DB 578 TATCGTTGAAGACGAGTTGGTAACAACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAG 637
QY 181 CTATGATGTTTTCGAAGCGACAGATGGCGCGAAATGATCAGATCCCTCTCTGAATATGA 240
DB 638 CTATGATGTTTTCGAAGCGACAGATGGCGCGAAATGATCAGATCCCTCTCTGAATATGA 697
QY 241 CATCAACTGCTGATCATGGATATCAATCTCCGCGGTAAGAACGCTCTCTGTTAGCGCG 300
DB 698 CATCAACTGCTGATCATGGATATCAATCTCCGCGGTAAGAACGCTCTCTGTTAGCGCG 757
QY 301 TGAACCTGCGAGCAGCGAATGTTGCGTTGATGTTCTCTGACT 343
DB 758 TGAACCTGCGAGCAGCGAATGTTGCGTTGATGTTCTCTGACT 800

RESULT 3
LOCUS BG834429
DEFINITION BG834429 114 bp mRNA linear EST 25-MAY-2001
ACCESSION BG834429
VERSION BG834429.1 GI:14199698
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
REFERENCE 1 (bases 1 to 114)
AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R.,
Quackenbush,J. and Keele,J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGTATGACCAT
BACKWARD: GTTTCCAGTCCAGC
Plate: 113 row: K column: 19
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..114
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
```

/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

ORIGIN

Query Match 12.3%; Score 114; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.9e-24;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 559 CCTGATGCGGAGAGTCAAGCTCCGCGAGCGAGTTCGCGCCATGCTTCACTTCTG 618
DB 1 CCTGATGCGGAGAGTCAAGCTCCGCGAGCGAGTTCGCGCCATGCTTCACTTCTG 60
QY 619 TGAACCCAGCGCAAAATTCAGTCCGCTGCTGAAGTCTGAAGAAATGACCGG 672
DB 61 TGAACCCAGCGCAAAATTCAGTCCGCTGCTGAAGTCTGAAGAAATGACCGG 114

RESULT 4
BG367933 841 bp mRNA linear EST 22-OCT-2001
LOCUS HVSME10014N08f Hordeum vulgare 20 DAP spike EST library HVCDA0010
DEFINITION (20 DAP) Hordeum vulgare subsp. vulgare cDNA clone HVSME10014N08f,
mRNA sequence.

ACCESSION BG367933
VERSION BG367933.1 GI:13257032
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 841)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D.,
Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,
Penton, R.D., Close, S.J., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex 20 DAP spike cDNA library
Unpublished (2001)

JOURNAL COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 285
Seq primer: AATTAACTCTACTAAAGGG
High quality sequence stop: 388.

FEATURES

source
1..841
Location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:12509"
/clone="HVSME10014N08f"
/tissue_type="20 DAP spike"
/lab_host="SOLR"
/clone_lib="Hordeum vulgare 20 DAP spike EST library
HVCDA0010 (20 DAP)"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Plants were grown in the greenhouse at the University of
California, Riverside (Penton, SJ Close, TJ Close). Whole
spikes with awns trimmed were collected at 20 DAP
(Penton). Total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give pBluescript
SK(-) cDNA phagemids in the TJ Close lab at the University
of California, Riverside (Choi). Phagemids were plated and

picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 12.1%; Score 112; DB 2; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.6e-23;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 816 AATCGCTTTACCAACCGTCAAAAAAAGCGCGCTTTTAGCGCGCTTTTATTTTCAAC 875
DB 1 AATCGCTTTACCAACCGTCAAAAAAAGCGCGCTTTTAGCGCGCTTTTATTTTCAAC 60
QY 876 CTTATTTCCAGATAGTAACCTCATCGCCGCTTGAACCTTCTTTACTGCGCTTT 927
DB 61 CTTATTTCCAGATAGTAACCTCATCGCTTGTGAACCTTCTTTACTGCGCTTT 112

RESULT 5

CNS00WTA 720 bp DNA linear GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC T12G15 of
DEFINITION TAMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.

ACCESSION AL094012
VERSION AL094012.1 GI:5295166
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (chale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 720)
Salanoubat, M., Choise, N., Artiguenave, F., Brottier, P., Wincker, P.,
Samson, D., Saurin, W., Weissenbach, J. and Quetier, F.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 720)
Genoscope.
AUTHORS Direct Submission

TITLE Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

FEATURES

source
1..720
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="T12G15"
/clone_lib="TAMU"
/ecotype="Columbia"
/note="end : T7"

ORIGIN

Query Match 11.0%; Score 102; DB 10; Length 720;
Best Local Similarity 54.9%; Pred. No. 2.3e-20;
Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 108 CCCGCACTTCTTATCGTTCAAGACGAGTTGGTTAAACGCAACGCTTGAAGATATTT 167
DB 350 CCATGAACGCTCTGCTGCTGCAAGACGACCGCGGATCGGTATACCTCTGCTACGGCT 409
QY 168 TCGAAGCGGAAGGCTATGATGTTTTTCGAAGCGACAGATGGCGCGGAATGCATCAGATCC 227

```

Db      410  TGAAGACCAATGCTTCGAGTGAAGACGCGCTCACGGCCAGAGGCGCTGATGCT 469
Qy      228  TCTCTGAATATGACATCAACCTGGTGATCATGATATCAATCTGCCGGTGAAGACGTC 287
Db      470  TTGGCGAAAAGCCCTCGAGTTCGTGATCTCGACATCGGCTGCGCCGACATGACCGCC 529
Qy      288  TTCTCTTAGCGGTCAACTGCGGAGCAGGCGAAATGTTGCGTTGATGTTCTGACTGCC 347
Db      530  TTGACGCTTGCCGGAACATCCCGAAAGCTCGACCGTCCGCGTGTCTTCTCTACCGTCC 589
Qy      348  GTGACACCAAGTCGATATAAATTTCTCGGCTCGAAATCGTGACATGACTACATCACCA 407
Db      590  GCGATGGCGAGTGACCGGATCTCTCGGCTCGAGCTCGCGGGGACGACTACGTGACCA 649
Qy      408  AACCTTTCAACCCCGGTGAATGACGATTCGTGCAACGAACTACTGTCCCGTACCATGA 467
Db      650  AGCCTTTAGCCCGCGGTGAATCGTGGCAGCATTCGTGCCATCTCTGCGCGCTCGACCA 709
Qy      468  ATCTCG 473
Db      710  GCCCG 715

```

```

RESULT 6
CL685986
LOCUS      CL685986
DEFINITION CL685986 768 bp DNA linear GSS 09-JUL-2004
            PRI0142c_H02_2 - PRI0142c.BR (768) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION CL685986
VERSION   CL685986.1 GI:50194404
KEYWORDS  GSS.
SOURCE    Pristionchus pacificus
           Pristionchus pacificus
           Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
           Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 768)
AUTHORS   Srivivasan,J., Otto,G.W., Kahlow,U., Geiseler,R. and Sommer,R.J.
TITLE     AppADB: an AcedB database for the nematode satellite organism
PUBMED    Pristionchus pacificus
JOURNAL   Nucleic Acids Res. 32 (1), D421-D422 (2004)
COMMENT   14681447
           Contact: Sommer RJ
           Evolutionary Biology
           Max-Planck-Institute for Developmental Biology
           Spemannstr. 37-39, Tuebingen D-72076, Germany
           Tel: 00497071601371
           Fax: 00497071601498
           Email: ralf.sommer@tuebingen.mpg.de
           This library was generated at Caltech, Pasadena, USA and end
           sequenced at Vancouver, Canada.
           Seq primer: r7
           Class: fosmid ends.

```

```

FEATURES             Location/Qualifiers
     source            1..768
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Mixed stage fosmid library of P. pacificus
                     var. California"
                     /note="Vector: pEpifos-5 Fosmid vector"

```

```

ORIGIN
Query Match      10.7%; Score 99.6; DB 10; Length 768;
Best Local Similarity 54.7%; Pred. No. 1.3e-19;
Matches 198; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

Qy      102  TCGACACCCCGGACATCTTATTCGTTGAAGACGAGTTGGTATACACGCAACGCTTGAAAA 161
Db      243  TGATGCCACATCACATTTGTTATTTGTTGAAGATGAGCCGGTTACCCAGGCGGATTAACAAT 302

```

```

Qy      162  GTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGACAGATGGCGCGGAATGCAATC 221
Db      303  CCTACTTCACTCAGAGGGGTATACCGTTTCCGTTACGGCAGCGGTCCCGGCTACGGG 362
Qy      222  AGATCCTCTCTGAAATATGACATCAACCTGGTGATCATGATATCAATCTGCCGGTAAAGA 281
Db      363  AAATATATGCAAGATCAGTCGGTAGATTTAATTCTGCTGGATATCAACTTACCCGATGAAA 422
Qy      282  ACGTCTTCTGTTAGCGGTGAACTGCGGAGCAGGCGAATGTTGCGTTGATGTTCTCTGA 341
Db      423  ATGCGCTGATGTTAAACCGCGCCCTCGGAGAAACGCTCAACCGTGGGATTTATCTGGTTA 482
Qy      342  CTGCGCGTGACAAAGAAATCGATAAAATTTCTCGGCTCGAAATCGTGACATGACTACA 401
Db      483  CCGACGCGAGCATCGGATTCACCGTATTTGGGCTGGAATGGCGGACACGATTAAG 542
Qy      402  TCACCAACCGTTCAACCCCGGTGAACCTGACGATTCGTGCAACGCAACCTACTGTCCCGTA 461
Db      543  TCACCAACCGCTGGAACCTGCGGAACTGGTGTAGTACGGGTGAAAAATCTGCTCTGGCGAA 602
Qy      462  CC 463
Db      603  TC 604

```

```

RESULT 7
BZ574214/c
LOCUS      BZ574214
DEFINITION BZ574214 1193 bp DNA linear GSS 17-DEC-2002
            msh2_3562.y3 msh Pseudomonas aeruginosa genomic clone msh2_3562,
            genomic survey sequence.
ACCESSION BZ574214
VERSION   BZ574214.1 GI:27209275
KEYWORDS  GSS.
SOURCE    Pseudomonas aeruginosa
           Pseudomonas aeruginosa
           Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
           Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1193)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol. (2002) In press
COMMENT    Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.

```

```

FEATURES             Location/Qualifiers
     source            1..1193
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="M5H"
                     /db_xref="taxon:287"
                     /clone_lib="msh2_3562"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."

```

```

ORIGIN
Query Match      10.1%; Score 93.2; DB 9; Length 1193;
Best Local Similarity 53.8%; Pred. No. 1.6e-17;
Matches 191; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

Qy      112  GCACATTTCTTATCGTTGAAGACGAGTTGGTAAACACGCAACAGTTGAAAAGTATTTTCA 171
Db      579  GCGCATCTGTTGATCGAAGACGATACGAACGCGCGAGTACTCTGAAGAGGGGCTCGG 520
Qy      172  AGCGGAAGGCTATGATGTTTTCGAAGCGACAGATGGCGCGGAATGCAATCGATCCTCTC 231

```

Db 519 CGAGTCGGGCTATGCGGTGCACTGGTGCAGACCGGCCGCGAGCGTCTCTACCTGGCGCT 460

Qy 232 TGAATATGACATCAACCTGGTGATCATGTGATCAATCTGCGGGGTAAAGACGGTCTTCT 291

Db 459 GGAGAACCGCTTCGACCTGGTGTCTCGACTGATGCTGCGCGCTGGACGGTTGGCA 400

Qy 292 GTTAGCGGTGAACCTGCGGACGAGCGAGATGTTGCGTTGATGTTCTCTGACTGCGCGGTGA 351

Db 399 GATCATGGAAGTGTGCGCAAGAAGCAGATGTCGGGTGCTCTTCTCCTCACGCGCGCGGA 340

Qy 352 CAACGAAGTCGATAAAATCTCGGCTCGAAATCGGTGCAGATGACTACATCACCACAC 411

Db 339 CCGCTGCAAGACCGTATCCGCGGCTCGACTGCGGTGCTGACGACTACTGCTGTAAC 280

Qy 412 GTTCAACCGCGTGAACCTGACGATTGTCGACGCAACCTACTGTCCCGTACCATG 466

Db 279 CTCTCTTTCACCGAGTTGCTCTGCTATCGTACCTCTGCTGCGCGCGGGTG 225

RESULT 8
CL652351/c
LOCUS
DEFINITION
CL652351 831 bp DNA linear GSS 09-JUL-2004
PRI0114c B11 - PRI0114c.B21 (831) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT
1 (bases 1 to 831)
Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

FEATURES
source
1. .831
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN
Query Match 10.0%; Score 92.6; DB 10; Length 831;
Best Local Similarity 54.6%; Pred. No. 2.3e-17;
Matches 185; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
Qy 102 TCAGACCCCGACATTTCTATGTTGAGACGAGTTGTTACACGCAACAGCTTGAAAA 161
Db 340 TGATGCCACATCACATTTGTTATGTTGAAGATGAGCCGGTTACCCAGCGCGATTACAA 281
Qy 162 GTATTTTCGAGCGGAAGCTATGATGTTTTCGAAGCGACAGATGCGCGGAAATGCAATC 221
Db 280 CCTACTTACTCAGAGGGGTATACCGTTTCGTTACGCGAGCGGTGCCGGGTACGGG 221
Qy 222 AGATCCTCTCTGAATATGACATCAACCTGGTGATCATGATATCAATCTGCGGGTAAGA 281

Db 220 AAATATTCAGAAATCAGTCGGTAGATTAAATTCGCTGGATATCAATTCACCGATGAAA 161

Qy 282 ACGGTCTTCTTAGCGGTGAATCGCGGACGAGCGGAATGTTGCGTTGATGTTCTCTGA 341

Db 160 ATGGCCTGATGTTAAACCGCGCCCTGCGAGAACGCTCAACGCTGGGGATTTCTGGTTA 101

Qy 342 CTGGCCGTGACACGAAGTCGATAAAATTCGCGCTCGAAATCGGTGCAGATGACTACA 401

Db 100 CCGGACGCGAGTCGGATTGACCGTATTGTTGGAAGGGAATGGCGCAGACGATTACG 41

Qy 402 TCACCAACCGTTCAACCCGCGTGAATGACGATTCGTG 440

Db 40 TCACCAACCGCTAAAGTGCAGAACTGCTAGTACGG 2

RESULT 9
BZ574107/c
LOCUS
DEFINITION
BZ574107 1018 bp DNA linear GSS 17-DEC-2002
msh2_3514.x1 msh Pseudomonas aeruginosa genomic clone msh2_3514,
genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 1018)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1. .1018
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="M5H"
/db_xref="taxon:287"
/clone_lib="msh2_3514"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN
Query Match 10.0%; Score 92.6; DB 9; Length 1018;
Best Local Similarity 53.8%; Pred. No. 2.4e-17;
Matches 191; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
Qy 112 GCACATTTCTATCGTTGAAGACGAGTTGGTTAAACGCAACAGTTGAAAAATTTTCGA 171
Db 560 GCGCATCTCGTGTGATCGAAGACGATACGAAGACGCGCGAGTACCTGGAAGAGGCGCTCGG 501
Qy 172 AGCGGAAGGCTATCATGTTTTCGAAGCGACAGATGGCGCGGAATGTCATCATCTCTC 231
Db 500 CGATCGGGCTATCGGGTCGACTGGTGCACACGCGCGCCGACGGTCTCTACCTGGCGCT 441
Qy 232 TGAATATGACATCAACCTGGTGATCATGGATATCAATCTGCGCGGTAAAGACGGTCTTCT 291
Db 440 GGAGAACCGCTACGACCTCGTGGTCTCTCGAGTGTCTGCCGCGCTGGACGGTTGGCA 381
Qy 292 GTTAGCGGTGAACCTGCGCGAGCGAGCGGAATGTTGCGTTGATGTTCTCTGACTGCGCGGTGA 351
Db 380 GATCATGGAAGTGTGCGCAAGAACGACGATGTGCGGTGCTCTTCTCCTCAGCGCCGGA 321

```

QY 352 CAACGAGTCGTAAATTTCTGGCCTCGAATCGGTGAGATGACTACATCACCACACC 411
Db 320 CAGCTGCAAGACCGTATCCGGCGCTCGAATCGGGTGTGACGACTACCTGGTGAACACC 261
QY 412 GTTCAACCGCGGTGAACGATTCGTGCGAGCAACCTACTGTCTCCCGTACCATTG 466
Db 260 CTTCTCCTCACCGAGTTGCTGCTGGTATCGTACCTGCTGCGCGCGGGGTG 206

RESULT 10
BZ559286
LOCUS
DEFINITION
  pacs2-164_1372.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION
  BZ559286
VERSION
  BZ559286.1 GI:27175379
KEYWORDS
  GSS.
SOURCE
  Pseudomonas aeruginosa
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
REFERENCE
  1 (bases 1 to 1159)
  Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
  Burns, J.L., Kaul, R. and Olsen, M.V.
  Whole-Genome-Sequence variation among multiple isolates of
  Pseudomonas aeruginosa library
JOURNAL
  J. Bacteriol. (2002) In press
COMMENT
  Contact: Chris K. Raymond
  Genome Center
  University of Washington
  Box 352145, Seattle, WA 98105-2145, USA
  Tel: 2062216954
  Fax: 2066857244
  Email: craymond@u.washington.edu
  Class: shotgun.
FEATURES
  source
  1..1159
  Location/Qualifiers
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone="pacs2-164_1372"
    /clone_lib="pacs2-164"
    /note="clinical isolate 2-164 Whole genomic shotgun
    library."
ORIGIN
  Query Match 9.7%; Score 89.8; DB 9; Length 1159;
  Best Local Similarity 55.2%; Pred. No. 1.9e-16;
  Matches 197; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 243 TCAACCTGCTGATCATGGATATCAATCTCGCGGGTAAGAACCGTCTTCTGTTAGCGGTG 302
Db 312 TCAGCTGTGTCTGCTGGACATCCGCTCGCCGCGCAAGGACGCGCTGACCTGACCCGG 371
QY 303 ACTGCGCGAGCAGGCGAATGTGTGATGTTCTGTACTGCGCGTGACACGAACTCG 362
Db 372 AGCTGCGGGTGGCTCCGAGGTGGGGATCATCTGATCACCGGGGCGCAACGACGAGATCG 431
QY 363 ATAAATTTCTCGGCTCGAATTCGGTGACAGTACTACATACCAACCGTTCAACCCGCG 422
Db 432 ATCGCATCTGCGCTGAGTGTGGGCGCGACGATTAAGTATCAAGCGCTGAACCCGCG 491
QY 423 GTGAATGACGATTCGTGCACGCAACCTACTCTCCGTTACCAATGTAATCTGGTACTGTGA 482
Db 492 GCGAATCTGTGTGCGGGCGAAGAAC--CTGATCCGCGGGTGGCCATGCCCCAGGCCA 548
QY 483 GCGAAGACGTGCTAGCTGTTGAAGCTACAAGTTCAATGGTTGGGAACTGGACATCAACA 542
Db 549 GCGCGGGCCCCCGCGGAGGCCCTCAGGCAGTTTGGCGACTGGCTGTGGACGCGGACC 608
QY 543 GCCGTTCTGTGATCGGCCCTGTATGGCGAGCAGTACAAGCTGCGCGGCGAGCGAGTTCC 599

```

```

Db 609 GGAGCCGCTGATGACACCGCGGCAACGAAACCTGCTCAACCATGGCGAATTCC 665

RESULT 11
BZ574473
LOCUS
DEFINITION
  meh2_3684.x1 meh Pseudomonas aeruginosa genomic clone meh2_3684,
  genomic survey sequence.
ACCESSION
  BZ574473
VERSION
  BZ574473.1 GI:27209534
KEYWORDS
  GSS.
SOURCE
  Pseudomonas aeruginosa
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
REFERENCE
  1 (bases 1 to 1575)
  Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
  Burns, J.L., Kaul, R. and Olsen, M.V.
  Whole-Genome-Sequence variation among multiple isolates of
  Pseudomonas aeruginosa library
JOURNAL
  J. Bacteriol. (2002) In press
COMMENT
  Contact: Chris K. Raymond
  Genome Center
  University of Washington
  Box 352145, Seattle, WA 98105-2145, USA
  Tel: 2062216954
  Fax: 2066857244
  Email: craymond@u.washington.edu
  Class: shotgun.
FEATURES
  source
  1..1575
  Location/Qualifiers
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="MSH"
    /db_xref="taxon:287"
    /clone="meh2_3684"
    /clone_lib="meh"
    /note="Environmental isolate. Whole genomic shotgun
    library."
ORIGIN
  Query Match 9.7%; Score 89.6; DB 9; Length 1575;
  Best Local Similarity 51.3%; Pred. No. 2.5e-16;
  Matches 234; Conservative 0; Mismatches 219; Indels 3; Gaps 1;
QY 345 GCCGTGACACGAAAGTCGATAAATTTCTCGGCCTCGAAATCGGTGCAGATGACTACATCA 404
Db 88 GCCCCACACGACGAGATCGATCGCATCGTCGCGCTGGAGTGGCGGCCGACGATTACGTGA 147
QY 405 CCMAACCGTTCAACCCCGGTGAACCTGACGATTTCTGTCACGCAACCTACTGTCCCGTACCA 464
Db 148 TCAAGCCGCTGAACCCCGCGCAACTGGTGTGCGGGGCGAAGAG--CTGATCCCGCGG 204
QY 465 TGAATCTCGGTACTGTCTAGCGAAGAACCTCGTAGCGTTTGAAGAAGTATCAAGTTTCAATGGTT 524
Db 205 TGGCCATGCCCCGAGCAGCCGCGCCCGCCGCGAGCCCTCAGGCAGTTTCGGGACT 264
QY 525 GGGAACTCGAATCAACAGCCGTTGTTGATCGGCCCTGTATGGCGAGCAGTACAAGCTGC 584
Db 265 GGTCTGTGAACGCGGACGCGGCGCCCTGTATCGACACGCGGGCAACGAAACCTGCTCA 324
QY 585 CGCGCAGCGAGTTCCGCGCCCATGCTTCACTTCTGTGAAACCCGACGAAATTCAGTCCC 644
Db 325 CCCATGGCGAATTCCAACTGCTGGCGCTTCTCTCGGCAACAGCAGGCAATACCTTAGCC 384
QY 645 GTGCTGAATCTGTGAAGAAATGACCGGCCGTGAGCTGAAACCGCACGACCGTACTGTAG 704
Db 385 GCGACCACTGATGGACAGATCCGCAACCGGAGTGGCTGCCAGCGATCGCTCCATCG 444
QY 705 ACGTGAAGATCGCGGTATTGTAACATTTTGAATCTACGCGCGATACGCGGGAATCA 764
Db 445 ACGTGTGTGCGCGCCTGCGTCGCAAGCTGCGCAGCAGACCGCGCCGAAACCGCAACTGA 504

```

```
Qy 765 TGGCCACCATTCACGGTGAAGTTATCGCTTCTGCG 800
Db |||||
505 TCATCACCATCCACGGCGCGGCTACCTGTTCCACG 540

RESULT 12
BZ551846 1029 bp DNA linear GSS 17-DEC-2002
LOCUS pacsl-60_3354.xl pacsl-60 Pseudomonas aeruginosa genomic clone
DEFINITION pacsl-60_3354, genomic survey sequence.
ACCESSION BZ551846
VERSION BZ551846.1 GI:27155556
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1029)
AUTHORS Burns,J.L., Raymond,C.K., Smith,E.B., Sims,E.E., Hastings,M.,
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
    source
        Location/Qualifiers
            1..1029
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="1-60"
                /db_xref="taxon:287"
                /clone="pacsl-60_3354"
                /clone_lib="pacsl-60"
                /note="clinical isolate 1-60 Whole genomic shotgun
                library."
ORIGIN
    Query Match 9.6%; Score 88.6; DB 9; Length 1029;
    Best Local Similarity 49.4%; Pred. No. 4.5e-16;
    Matches 283; Conservative 0; Mismatches 286; Indels 4; Gaps 2;

Qy 243 TCAACCTGGTATCATGATATCAATCTGCCGGGTAAGACGGTCTTCTGTAGCGGTG 302
Db |||||
202 TCGAGCTGGTGTCTGTGACATCCGCTGCCCGCAAGACGGCTGACCCCTGACCCGCG 261

Qy 303 AACTGCGGACGACGGCGAATGTGGTGTGATGTTCTGACTGGCGGTGACACGAGTCG 362
Db |||||
262 AGCTGCGGGTGGCTCCGAGGTGGGATCATCTTGATCACCGGGCGCAACGAGATCG 321

Qy 363 ATAAATTTCTGGCTTCGAAATCGGTGCAGATGACTACATCAACAAACGTTCAACCCGC 422
Db |||||
322 ATGCATCGTGGCTTGGATGCGGCGCGACGATTAGTGATCAAGCGGTGACCCGC 381

Qy 423 GTGAATGACGATTCGTGCAGCAACCTACTGTGCCGTAACATGAATCTGGGTACTGTCA 482
Db |||||
382 GCGAATCGTGTGCGGGCGAAGAAC---CTGATCCGCGGTGGCCATGCCAGGCCA 438

Qy 493 GCGAAGACGTCGTAGCGTTGAAGCTACAGTTCAATGGTTGGAACTGGACATCAACA 542
Db |||||
439 GCGCGGGCCCGCGCGGCGAGCCCTTAAGCGAGTTGCGGCACTGGCTGTGACGCGGACC 498

Qy 543 GCGGTTCGTGTATCGCGCTGATGCGGACGAGTACAAGCTGCGCGCAG-CGAGTTTCGC 601
Db |||||
499 GCGCGCGCTGTATGACACACGCGGCAACAAACCTGCTCACCATCGNGCAATTCA 558

Qy 602 GCCATGCTTCACTTCTGTGAATAACCCAGGCAAAATTCAGTCCCGTGTGAATGCTGAAG 661

Qy 559 TTGCTCGGCGCTTTTCTGCGCAACAGCGGGGCATACCTCGAGCCGACAGTTCATGGAC 618
Db |||||
662 AAAATGACCGCGCGTGAAGCTGAACCGCACGACGCTACTGTAGAGCTGACCATCGCGGT 721
Db |||||
619 CAGATCCGCAACCGGAGTGGCTGCCAGGATCGCTCTTCTGCTGGCGCGGCGCCCT 678
Qy 722 ATTCGTAACATTTTCAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACCGGT 781
Db |||||
679 GCGTCGACGCTGTGCGAGCGACCGCGCAACCGCACTTGATCATACCATCACGNGCGGT 738
Qy 782 GAAGGTTATCGCTTCTGCGGTGATCTGGAAGAT 814
Db |||||
739 ACTTGTACCGCGCGCGCAGCGACGCTGAGCAT 771

RESULT 13
CL944656 732 bp DNA linear GSS 21-SEP-2004
LOCUS OsIFS8005271 Oryza sativa Express Library Oryza sativa (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL944656
VERSION CL944656.1 GI:52356665
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 732)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
JOURNAL its comparison to Arabidopsis
COMMENT Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
    source
        Location/Qualifiers
            1..732
                /organism="Oryza sativa (indica cultivar-group)"
                /mol_type="genomic DNA"
                /db_xref="taxon:3946"
                /clone_lib="Oryza sativa Express Library"
                /note="Oryza sativa exon trapped genomic sequences "
ORIGIN
    Query Match 9.1%; Score 84.8; DB 10; Length 732;
    Best Local Similarity 47.0%; Pred. No. 6.4e-15;
    Matches 331; Conservative 0; Mismatches 367; Indels 6; Gaps 2;

Qy 96 CAACATGACAGCCCGCAGCATTTCTATCGTTGAAGCAGAGTTGGTACACGACACAGT 155
Db |||||
11 CAACCAACCGTACCGACAAGATCTCTGTTGGGACGACGCGCGCATCCGTGATTTGC 70

Qy 156 TGAAGAATGATTTTCGAAGCGGAAGCTATGATGTTTTCGAAGCGACAGATCGCGGAAA 215
Db |||||
71 TCGCCCGCTATCTGACCCAGGAGGCTTCGAAGTCATGTTGCGGAGACGCGCAGGCGC 130

Qy 216 TGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGAATCATGATATCAATCTGCCGG 275
Db |||||
131 TCAACCGCATCTCTGCTGCGGAGACGGTTCGAACATGATCGTTCTGATCTCATGATGCCG 190

Qy 276 GTAGAAGCGGTCTTCTGTAGCGGTGACCTGACGCGGAGC---AGCGAATGTTGGTTGA 332
Db |||||
191 GCGAGGACGGGCTCTCCATCTGCCCGCGCTGCGCGGCGCAACGACCGCACCCCGATCA 250
```



```
QY 333 TGTTCCTGCTGGCGGTGACAGGAGTCGATAAAATTTCTGGCCTCGAAATCGGTGCAG 392
Db 251 TATGCTACACCCCAAGCGGAGGACGTGGATCGCATCGTGGGTGGAAAGTGGTGCCG 310
QY 393 ATGACTATACATCAACAAACCGTTCAACCGCGTGAACATGACGATTCTGTCAGCGCAACCTAC 452
Db 311 ACGACTACTGGGCAAGCCCTTCATCCCGGAGCTGCTGGCGGCATCCAGCCGTGC 370
QY 453 TGTCCCGTACCATGAATCTGGGTACTGTACAGCAAGAAACGTGTAGCTGTGAAGCTACA 512
Db 371 TGCAGCGCGCCCGCGCAGGAGGACACCGGCGCCCTCGGGCGACAAACGAGGTCTGA 430
QY 513 AGTTCAATGGTGGGNACTGGACATCAACAGCCGTTCTGTGATCGGCCCTGATGCGGAGC 572
Db 431 CTTTCGGGCCCTTCACCTTCGACCTGGGCAACGCGCCCTGCGAAGA---ACGCGAGG 487
QY 573 AGTACAAGCTGCGCGGACGAGTTCCGCGCCATGCTTCACTTCTGTGAATAACCCAGGCA 632
Db 488 AGCTGCCCTCACACCGCGGAATTCGCCATGCTCAAGGCCCTGTGTGCGCCACCGCGCC 547
QY 633 AAATTCAGTCCGCTGTAATCTGTGTAAGAAATGACCGCGCTGAGCTGAAACCGCACG 692
Db 548 AGCCGCTGTGCGCGCAAAAGCTGGCCCTGCTGGCCCGCGCGCGAGTTTCGAGCCCTTCG 607
QY 693 ACCGTACTGTAGACGTGACATCCCGCTGATTCGTGAACATTTCAATCTACCGCGATA 752
Db 608 ACCGAGCCGTGACGTGACAGTCTCGCGCTCGCGCAAGCTGTGAGTGATGATCGCGCG 667
QY 753 CGCCGGAATCATCGCCACCATTCACGGTGAAGTTATCGCTTC 796
Db 668 CTCCCCGCTACATCCAGACCGTCTGGGCGTGGGTACGTGTC 711
```

```
RESULT 14
CW158500/c 665 bp DNA linear GSS 29-OCT-2004
LOCUS 104_564_11148854_148_36401_059 Sorghum methylation filtered library
DEFINITION (LibID: 104) Sorghum bicolor genomic clone 11148854, genomic survey
sequence.
ACCESSION CW158500
VERSION CW158500.1 GI:54851047
KEYWORDS GSS.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 665)
AUTHORS Bedell J.A., Rudiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,
Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,
McMenamy, J., Smith, M., Holsman, H., Roe, B.A., Wiley, G., Korff, I.F.,
Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddalo, J.A. and
Martienssen, R.A.
TITLE Sorghum genome sequencing by methylation filtration
JOURNAL PLoS Biol. 3 (1), e13 (2005)
PUBMED 15660154
COMMENT Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 564 row: f column: 14
Seq primer: SWfor Forward
Class: methylation filtered
High quality sequence stop: 665.
```

```
FEATURES
source
1..665
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="Atx623"
/db_xref="taxon:4558"
```

```
/clone="11148854"
/clone_lib="Sorghum methylation filtered library (LibID:
104)"
/organism="Sorghum bicolor"; Vector: pBSK(-); Site: 1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBSK(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."
```

```
ORIGIN
Query Match 8.9%; Score 82.8; DB 10; Length 665;
Best Local Similarity 52.3%; Pred. No. 2.7e-14;
Matches 183; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 110 CCGCACATCTTATTCGTTGAAGACGAGTTGTGAACACGCAACACGTTGAAAGATTTTC 169
Db 471 CCCACATCTCTGTCGTCGATGACGACACCGCCCTGCGCAACTGCTCGCAAGTATCTC 412
QY 170 GAAGCGGAGGCTATGATGTTTTCGAGCGACAGATGCGCGGAAATCATCATCATCTC 229
Db 411 GCCGCGCAACGCTTCTGTCGCGCAAGGACGCGCGGAGCGCGCGCCGCTG 352
QY 230 TCTGAATATGACATCAACCTGTCGTCGATGATGATGATGATGTCGCGGGTAAAGACGCTT 289
Db 351 GCCGGCTCGCTTCGACCTGCTGTCGTCGATGATGATGATGATGATGATGATGATGATG 292
QY 290 CTGTTAGCGCGTGAACGTCGCGGAGGAGCGAATGTTGGTTGATGTTCTGATGCGCGT 349
Db 291 GAGCTGACCGAATCGCTGCGCGGAGCGACCTGCGCATCTGCTGCTGACCGCGG 232
QY 350 GACACGAGGTCGATAAAATTTCTGGCTCGCAATCGTGCAGATGATGATGATGATGATG 409
Db 231 GCGGAGCGGACGACCGCATCGCGGTTGGAGCGCGCGCGCGAGCATCTCTGCTGCGCAG 172
QY 410 CCGTTCAACCGCGTGAACATGACGATTCGTCGACGCAACCTACTGTCCTCCG 459
Db 171 CCTTCAACCGCGGAGCTCTGTCGCGCATCAACTCGATCTGCGCG 122
```

```
RESULT 15
BZ564167/c 1189 bp DNA linear GSS 17-DEC-2002
LOCUS pac2-164_4617_xl pac2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pac2-164_4617, genomic survey sequence.
ACCESSION BZ564167
VERSION BZ564167.1 GI:27189394
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1189)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
```

```
FEATURES
source
1..1189
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_4617"
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2006, 13:37:39 ; Search time 215 Seconds
(without alignments)
7664.184 Million cell updates/sec

Title: US-10-613-990A-31

Perfect score: 927

Sequence: 1 gtcattgtacgcgatcatg.....gtacttttactggcttt 927

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/FCUS COMB.seq.*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577.2	62.3	726	3	US-09-489-039A-1812
2	450.8	48.6	726	3	US-09-543-681A-2033
3	319.4	34.5	1830121	3	US-09-557-884-1
4	319.4	34.5	1830121	3	US-09-643-990A-1
5	319.4	34.5	1830121	3	US-10-158-865-1
6	126	13.6	837	3	US-09-252-991A-16417
7	126	13.6	2430	3	US-09-252-991A-16171
8	126	13.6	3675	3	US-09-252-991A-16309
9	120.6	13.0	990	3	US-09-252-991A-11517
10	120.6	13.0	2166	3	US-09-252-991A-11722
11	113.8	12.3	2004	3	US-09-252-991A-11865
12	107.4	11.6	1018	3	US-09-418-980-1
13	98.6	10.6	729	3	US-09-634-238-133
14	97.2	10.5	762	3	US-09-252-991A-6307
15	96.2	10.4	705	3	US-09-134-001C-226
16	94.6	10.2	702	3	US-09-710-279-479
17	94.6	10.2	3389	3	US-09-710-279-3914
18	92.6	10.0	792	3	US-09-252-991A-10543
19	92.6	10.0	804	3	US-09-252-991A-10323
20	91.4	9.9	699	3	US-09-925-637-45
21	91.4	9.9	702	3	US-09-082-077-2
22	91.4	9.9	3731	3	US-09-082-077-1
23	91.4	9.9	9425	3	US-08-956-171E-87
24	91.4	9.9	9425	3	US-08-781-986A-87

ALIGNMENTS

RESULT 1

US-09-489-039A-1812
; Sequence 1812, Application US/09489039A
; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 1812

; LENGTH: 726

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-1812

Query Match	62.3%	Score 577.2;	DB 3;	Length 726;
Best Local Similarity	87.2%	Pred. No. 3.3e-190;		
Matches 633;	Conservative	0;	Mismatches 93;	Indels 0; Gaps 0;
Qy	92	GTAGCAAAACATGCGAGACCCCGCACATTTCTTATCGTTGAAGACGAGTTGGTAACACGCAAC	151	Sequence 3722, Ap
Db	1	GTAGCAAAACATGCGAGACCCCGCACATTTCTTATCGTTGAAGACGAGTTGGTAACACGCAAC	60	Sequence 3092, Ap
Qy	152	ACGTTGAAAAGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGGACAGATGCGCGG	211	Sequence 2878, Ap
Db	61	ACGTTGAAAAGTATTTTCGAAGCGGAAGGTTACGATGATTTTCGAAGCGGACGATGCGCGG	120	Sequence 2990, Ap
Qy	212	GAATGCGATCAGATCCTCTCTGATATGACATCAACCTGGTGATCATGATATCATCTG	271	Sequence 1321, Ap
Db	121	GAATGCGATCAGATCCTCTCTGATATGACATCAACCTGGTGATCATGATATCAACCTG	180	Sequence 14821, A
Qy	272	CGGGTAAGAACGGTCTTCTGTTAGCGGCTGAACCTCGCGGACGAGCGAATGTTGGTGTG	331	Sequence 14543, A
Db	181	CGGGTAAGAACGGTCTTCTGTTAGCGGCTGAACCTCGCGGACGAGCGAATGTTGGTGTG	240	Sequence 12, Appl
Qy	332	ATGTTCTTGACTGGCGGTGACAAACGAAGTCGATAAAATTTCTCGGCTCGAAATCGGTGCA	391	Sequence 1, Appl
Db	241	ATGTTCTTAAACCGGCGCGGACAAACGAAGTGGATAAGATCTTTGGCTCGAAATCGGCGCT	300	Sequence 5924, Ap
Qy	392	GATGACTACATCACCACAAACCGTTCAACCGCGGTGAACGATGCGATGCGTGCAGCAACCTA	451	Sequence 5003, Ap
Db	301	GACGACTATATCATAAACCCGTTTAAACCGCGGTGAACCTGACTATCGGCGCGGCAACCTG	360	Sequence 1923, Ap

QY 452 CTGTCCTCCGTACCATGAATCTGGTACTGTCTCAGCGAAGAACTGCTAGCGTTGAAAGCTAC 511
DB 361 CTCTCCCGTACCATGAACCTCGGTACCGTGAGCGAAGAGCGCTCGCAGCGTGGAAGCTAC 420
QY 512 AAGTTCAATGGTTGGGAATGAGACATCAACAGCGGTTCTGTTGATCGGCCCTGTAGTGGCGAG 571
DB 421 AAGTTCAACGGCTGGGAATCTGATATCAACAGCCGTTCTCCTGGTTAGCCGNAACGGCGAA 480
QY 572 CAGTACAAGCTGCCCGCAGCGAGTTCCGCGCAATGCTTCACTTCTGTGAAACCCAGGC 631
DB 481 CAGTACAAGCTGCCCGCAGCGAGTTCCGCGCGATGCTGCACTTCTGCGAAGAACCCGGGC 540
QY 632 AAAATTTCAGTCCCGTCTGAACTGCTGAAGAAATGACGGCCGCTGAGCTGAACCGCAC 691
DB 541 AAAATTTCAGTCCCGTCTGAACTGCTGAAGAAATGACGGCTCGCGAGCTGAAGCCGAT 600
QY 692 GACCGTACTGTAGAGCTGACGATCCGCGTATTCGTAAACATTTTGAATCTACGCCGAT 751
DB 601 GACCGTACTGTAGAGCTGACGATCCGCGTATTCGTAAACATTTTGAATCTACGCCGAT 660
QY 752 AGCCGGAATATCATCCCAACCATTCACCGTGAAGGTTATCGCTTCTGCGGTGATCTGAA 811
DB 661 ACCCGGAATATCATCCCAACCATTCACCGTGAAGGTTATCGCTTCTGCGGTGATCTGCA 720
QY 812 GATTAA 817
DB 721 GAATAA 726

RESULT 2

US-09-543-681A-2033

; Sequence 2033, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 2033

; TYPE: DNA

; LENGTH: 726

; ORGANISM: Proteus mirabilis

US-09-543-681A-2033

Query Match 48.6%; Score 450.8; DB 3; Length 726;
Best Local Similarity 76.3%; Pred. No. 3.8e-146;
Matches 554; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 92 GTAGCAAAATATGCAACCGCCGACATCTTATCGTTGAAGCGAGTGTGAACAGCAAC 151
DB 1 GTAGCAAAATATGCAACCGCCGACATCTTATCGTTGAAGCGAGTGTGAACAGCAAC 60
QY 152 AGCTTTGAAAGTATTTTGAAGCGGAGCTATGATGTTTTTGAAGCGACAGATCGCGG 211
DB 61 ACCCTGAAAGCATATTCGAAGCTGAAGGTTATCGTACACGAAGCACTGATGGCAAC 120
QY 212 GAAATGCATCAGATCTCTCTGAAATATGACATCAACCTGGTGATCATGTATCAATCTG 271
DB 121 GAGATGCATATATTTCTGTCGACCATGATATCAATCTGGTCAATATGATATTAATCTT 180
QY 272 CCGGGTAAGAACGGTCTTCTGTTAGCGGTGAACCTGCGGACGACGAGTGTTCGGTTG 331
DB 181 CTGGTAAGAAATGGTCTTCTATAGCCCGTGAATACGTGAACAGGTAGTGTTCATTA 240
QY 332 ATGTTCTGACCTGGCGTGACAAACGAAGTCGATAAAATTTCTCGGCCCTCGAAATCGGTGCA 391
DB 241 ATGTTCTTAACAGGTCTGTGATATGAAGTTGATAAAATCTTAGGCCCTTGAAATTCGTGCC 300

QY 392 GATGACTACATCACCBAACCGTTTCAACCGCGGTGAACCTGACGATTCGTGACGCAACCTA 451
DB 301 GATGATTACATCACTAAACCAATTTAATCCTCGTGAAATTAACCTATCGGTGCTCGTAACCTA 360
QY 452 CTGTCCTCCGTACCATGAATCTGGTACTGTCTCAGCGAAGAACTGCTAGCGTTGAAAGCTAC 511
DB 361 TTGTCAACGCACTATGAATTTAGCGAATGGCACAGAGAGCGTCTGTTAGTTGAAAGCTAT 420
QY 512 AAGTTCAATGGTTGGGAATGAGACATCAACAGCGTTCGTTGATCGGCCCTGTAGTGGCGAG 571
DB 421 AAATTTAATGGTTGGGAGCTAGATATTAATAAGTGGCTCTCTATTATAGCCCTACAGGTGAA 480
QY 572 CAGTACAAGCTGCCCGCAGCGAGTTCCGCGCCATGCTTCACTTCTGTGAAACCCAGGC 631
DB 481 CAGTATAAATTTACCTCGTAGTGAGTTTCGTGCGGATGTTACATTTCTGCGAAGAACCCAGAA 540
QY 632 AAAATTTCAGTCCCGTCTGAACTGCTGAAGAAATGACCGGCCGCTGAGCTGAAACCCGAC 691
DB 541 AAAATCCAACTCGTGAGATTAATGAAAGAAATGACGGGTCTGTAATTAACCTCAT 600
QY 692 GACCGTACTGTAGAGCTGACGATCCGCGTATTCGTAAACATTTTGAATCTACGCCGAT 751
DB 601 GATCGTACTGTAGAGCTTACCATTCGTGATTCGTAAACACTTTGAAATCAACCCCTGAT 660
QY 752 ACGCCGGAATATCATCCCAACCATTCACGGTGAAGGTTATCGCTTCTGCGGTGATCTGAA 811
DB 661 ACACCTGAGATTAATCGCCACTATCCATGCGTGAAGGTTATCGTTTCTGTTGGTGAATTTAGAC 720
QY 812 GATTAA 817
DB 721 GAGTGA 726

RESULT 3

US-09-557-884-1/c

; Sequence 1, Application US/09557884

; Patent No. 6506581

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments

Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match      34.5%; Score 319.4; DB 3; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.2e-98;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTAGTAGCAACATCGACAGCCCGCACATCTTATCGTTGAAGCAGGTTGGTAAC 144
DB 936640 AGTAAAGATGAGAAATGACTACTCCAAATCTCGTTGTGAAGATGAATTTGTCCAC 936581

QY 145 ACGCAACACGTTGAAAGTATTTTCGAAGCGAAGCTATGATGTTTTCGAAGCGACAGA 204
DB 936580 TCGAATACGCTTAAGGATTTTGAAGCGAAGATATGATGTTTGAAGCGAATA 936521

QY 205 TGGCGCGGAAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGGATAT 264
DB 936520 TGGTGTGAATGCATCATATATTGGCAATCATATATTAATTTGGTTGTGATGATAT 936461

QY 265 CAATCTGCGGTAAGACGGTCTTCTTGAAGCGGTGAATCGCGGAGCAGGCGAATGT 324
DB 936460 TAAATTTACCGCAAAAACGGCTTATTATTGGCAAGAGAACTCCGTTGAAGAAATTAAGCTT 936401

QY 325 TGGCTTGATGTTCTGACTGGCGGTGACAAAGAGTGCATATAAATTCGGCTCGAAT 384
DB 936400 ACCTCTTATTTTAACTGGTGGAGATTAAGTGGATATAAATTTAGGTTTGGAAAT 936341

QY 385 CGGTGCAGATGACTACATCACCACACCGTTCAACCCGCGGTGAATGACGATTCGTGCACG 444
DB 936340 TGGTCCGATGACTATTGACTAAGCTTTTAAACCTAGAGAACTGACTATCCGCGCAG 936281

QY 445 CAACCTACTGTCCTGACATGAATCTGGGTACTCTGACGGAAGAACTCGTAGGTTGA 504
DB 936280 TAACTTATGTCATGCTGCAATG---CCGCATCAAGAAAAAGAAATACATTTGTCGAGA 936224

QY 505 AAGCTACAAAGTTCAATGTTGGGAAGTGCAGATCAACAGCCGTTGTTGATCGGCCCTGA 564
DB 936223 ATTCTATCGTTTAAAGTTGGAATTAAGCTTAATAGCCATGAATTAATTAACACAGA 936164

QY 565 TGGCGAGCAGTACAAGCTCGCGCGCAGCGAGTTTCGCGCCATGCTTCACTTCTGTGAAAA 624
DB 936163 AGGACAAGATTCAAACTCTCTCGCAGTGAATTTCTGTCAGATTTACATTTCTGTGAAA 936104

QY 625 CCCAGGCAAAATCGATCGCGGTGCTGAATCTGCTGAAGAAATGACCGCGCGTGAAGTAA 684
DB 936103 TCCAGGAAAAATGCAACCGCGCGAAGAAATTCCTGAAAAAATGACCGGACGAGAGTTAAA 936044

QY 685 ACCGACACCGTACTGTAGAGTGCAGATCCGCGGTATTCGTAAACATTTTCGATCTAC 744
DB 936043 ACCTCAGGATCGTACCGTAGATGTACAAATTCGAGTATCGAATAACATTTGAAGATCA 935984

QY 745 GCGGATACGCGGAAATCATCGCCACCAATTCACGGTGAAGGTATCGCTTCTCGCGTGA 804
DB 935983 TCCCAATCTCAATATCATATGACTATACATGAGAGGCTATCGTTTTGTGGAGA 935924

QY 805 TCTGGAAGA 813
DB 935923 TATTGAATA 935915
```

RESULT 4
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter

```

;
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PBI86P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
```

```

Query Match      34.5%; Score 319.4; DB 3; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.2e-98;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTAGTAGCAACATCGACAGCCCGCACATCTTATCGTTGAAGCAGGTTGGTAAC 144
DB 936640 AGTAAAGATGAGAAATGACTACTCCAAATCTCGTTGTGAAGATGAATTTGTCCAC 936581

QY 145 ACGCAACACGTTGAAAGTATTTTCGAAGCGAAGGCTATGATGTTTTCGAAGCGACAGA 204
DB 936580 TCGAATACGCTTAAGGATTTTGAAGCGAAGATATGATGTTTGAAGCGAATA 936521

QY 205 TGGCGCGGAAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGGATAT 264
DB 936520 TGGTGTGAATGCATCATATATTGGCAATCATATATTAATTTGGTTGTGATGATAT 936461

QY 265 CAATCTGCGGTAAGACGGTCTTCTTGAAGCGGTGAATCGCGGAGCAGGCGAATGT 324
DB 936460 TAAATTTACCGCAAAAACGGCTTATTATTGGCAAGAGAACTCCGTTGAAGAAATTAAGCTT 936401

QY 325 TGGCTTGATGTTCTGACTGGCGGTGACAAAGAGTGCATATAAATTCGGCTCGAAT 384
DB 936400 ACCTCTTATTTTAACTGGTGGAGATTAAGTGGATATAAATTTAGGTTTGGAAAT 936341

QY 385 CGGTGCAGATGACTACATCACCACACCGTTCAACCCGCGGTGAATGACGATTCGTGCACG 444
DB 936340 TGGTCCGATGACTATTGACTAAGCTTTTAAACCTAGAGAACTGACTATCCGCGCAG 936281

QY 445 CAACCTACTGTCCTGACATGAATCTGGGTACTCTGACGGAAGAACTCGTAGGTTGA 504
DB 936280 TAACTTATGTCATGCTGCAATG---CCGCATCAAGAAAAAGAAATACATTTGTCGAGA 936224
```



```
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature

Query Match      34.5%; Score 319.4; DB 3; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.2e-98;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTTAGGTAGCAAAACATGACAGACCCCGCACATCTTTATCGTTGGAAGACGAGTTGGTAAC 144
DB 936640 AGTATAAGGATGGAAATGACTACTCCAAAATTTCTCGTTGTTGAAGATGAAATTTGTCCAC 936581

QY 145 ACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGACAGA 204
DB 936580 TCGAAATACGCTTAAAGGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCGAATAA 936521

QY 205 TGGCGCGGAAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATCGATAT 264
DB 936520 TGGTGTGAAATGCATCATATATTGGCAATCATATATTAATTTGGTTGTGATGATAT 936461

QY 265 CAATCTCGCGGTAAAGAACGGTCTTCTGTAGCGGTGAATCGCGCGAGCAGGCGGAATGT 324
DB 936460 TAATTTACGAGCAAAAACGGCTTATTATTTCGCAAGAGAACTCCGTGAAGAAATTAAGCTT 936401

QY 325 TGGTGTGATGTTCTGACTGCGCGTGACCAAGAGTCGATATAAATTTTCGCGCTCGAAAT 384
DB 936400 ACCTCTTATTTTTTAACTGTCGAGATAAAGAGTGGATAAATTTTAGGTTTGGAAAT 936341

QY 385 CGGTGCGAGTACTACATCACAAAACGGTTCACCCCGGTCAACTGACGATTCGTGCGACG 444
DB 936340 TGGTGGCGATGACTATTGACTAAGCCTTTTAAACCTAGAGAACTGACTATCCGCGCACG 936281

QY 445 CAACCTACTGTCCCGGTACCATGAATCTGGGTACTGTCTCAGCGAAGAAACGTCGTAGCGTTGA 504
DB 936280 TAATCTATTGCATCGTCAATG---CCGCATCAAGAAAAAGAAATACATTTGTCGAGA 936224

QY 505 AAGCTACAAGTTCAATGTTGGGAACTGGGACATCAACAGCCGTTGTTGATCGGCGCTTGA 564
DB 936223 ATTCTATCGTTTTAATGTTGGAAATTAGACCTAAATAGCCACAGTTTTAATTTACACCA 936164

QY 565 TGGCGAGCAGTACAAGCTCCGCGCGCAGCGAGTTCGCGCCACTGCTTCACTTCTGTGAAA 624
DB 936163 AGGACAAGAAATTCAAACTTCTCGCAGTGAATTTCTCGCAATGTTTACATTTCTGTGAAA 936104

QY 625 CCCAGGCAAAAATTCAGTCCCGTGTGAACTGCTGAAGAAAAATGACCGGCCGTGAGCTGAA 684
DB 936103 TCCAGGAAAAATTCGAAACGCGCGAAGAAATTCCTGAAAAAATAGCCGACGAGAGTTAAA 936044

QY 685 ACCGACGACCGTACTGTAGACGTGACGATCCGCGGTATTCGTAAACATTTTCGAATCTAC 744
DB 936043 ACCTCAGGATCGTACCGGTAGATGTCAAAATTCGACGTATCAGAAAAACATTTTGAAGATCA 935984

QY 745 GCCGATACGCGGAAATCATCGCCACCATTCAGGTGAGGTTATCGCTTCTCGGGTGA 804
DB 935983 TCCCAATACTCCAAATATCATATGACTATATCATATGAGAGAGCTATCGTTTTTGTGGAGA 935924

QY 805 TCTGGAAGA 813
DB 935923 TATTGAATA 935915
```

RESULT 6

```
US-09-252-991A-16417
; Sequence 16417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

; PRIOR FILING DATE: 1998-02-18

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11722

Query Match 13.6%; Score 126; DB 3; Length 3675;
Best Local Similarity 52.9%; Pred. No. 1.7e-32;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;
QY 243 TCAACCTGGTATCATGATATCAATCTGCCGGGTAAGAACGGTCTTCTGTTAGCGGTG 302
DB 72 TCGAGCTGGTGTCTGGACATCCGCTGCCGGCAAGGACGGCTGACCCCTGACCCGCG 131
QY 303 AACTCGGAGCAGCGGAATGTTGGTTGATGTTCTGACTGCGCGTGAACAAGTGG 362
DB 132 AGCTCGGGTGGCTCCGAGGTGGGGATCATCTGATCACCGGGGCAACGACGAGATCG 191
QY 363 ATAAAAATTCCTCGGCTCGAAATCGGTGCGAGTACTACATCAACCAACCGTTCAACCCGC 422
DB 192 ATCGCATGCTCGGCTGAGTGGCGCGGACGATTAATGATCAAGCGCTGAACCCGC 251
QY 423 GTGAATGACGATTCGTGCAACCACTACTGTCCCGTACCATGAATCTGGGTACTGTCA 482
DB 252 GCGAACTGTGTGCGGGG---GAAGAATCTGATCCGCGGGTGGCCATGCCAGGCCA 308
QY 483 GCGAAGACGTGCTAGCTGTAAGCTAAGTTCATGTTGGTGGAACTGGACATCAACA 542
DB 309 GCGCGGGCCCGCGGAGGCGCTCAGGCACTTCGGGACTGGCTGTGACGCGCGACC 368
QY 543 GCGCTTGTGATCGGCTGATGCGAGCAGTACAACTGCGCGCAGCGAGTTCCGCG 602
DB 369 GCGCGGCTGATGACACACGCGGGCAAGAAACCTGCTACCATGCGGAATCCAAAT 428
QY 603 CCATGCTTCACTTCTGTGAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGTGAAGA 662
DB 429 TGCTCGGCGCTTCTCGGCAACAGCGGCATACCTGAGCGCGCACTGATGGACC 488
QY 663 AAATGACCGCGTGTAGCTGAACCGCAGCAGTACTGTGATGATGATGATGATGATG 722
DB 489 AGATCGCAACCGAGTGGCTGCGCAGCGATCGCTCCATGATGATGATGATGATG 548
QY 723 TTCTGAAACATTTGAAATCTAGCGGATAGCGGAAATCATCGCCACCATTCACGTTG 782
DB 549 TCGTGTGCAAGTGGCGGAGGACCGCGGAAACCGCACTGATCATCATTCACCGCG 608
QY 783 AAGTTATCGTCTTGGCG 800
DB 609 CCGGCTACCTGTTCCCG 626

RESULT 9

US-09-252-991A-11517/c
; Sequence 11517, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11517
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11517

Query Match 13.0%; Score 120.6; DB 3; Length 990;
Best Local Similarity 49.2%; Pred. No. 5.6e-31;
Matches 349; Conservative 0; Mismatches 354; Indels 6; Gaps 1;

QY 95 GCAAAACATGCAAGCCCGCACATCTTATCTGTTGAAGACGAGTTGGTAACACGCAACACG 154
DB 747 GCAATGGAGCATGTGATCACAATCTGATCTGTCGACGATGACCGCGAGATTCGCGAGCTG 688
QY 155 TTGAAAAGTATTTTGAAGCGGAAGGCTATGATGTTTTCGAAGCGACAGATGGCGCGGAA 214
DB 687 GTAGGCAACTACTTGAAGAAGACGGCTGCGCACCAACCATCTGTCGCGCGATGGCGCGGAG 628
QY 215 ATGCATCAGATCTCTCTGAAATATGACATCAACCTGGTGATCATGATATCAATCTGCGG 274
DB 627 ATGCGCGGTTCTTGAAGCCCAACAGGTGACCTGATCTGCTCTGCAATCATGATGCC 568
QY 275 GGTAAAGACGCTTCTTGTAGCGCGTGAACCTGGCGG-----AGCAGCGCAATGTTGCG 328
DB 567 GCGACGAGCGCTGCTGCTGCGCGGAGCTGCGCGTGGCGGCGCAACAGCCACGCGCG 508
QY 329 TTGATGTTCTTGAAGTGGCGGTGACAAAGAAATGATGATGATGATGATGATGATGATG 388
DB 507 GTGCTGATGCTTCAACCGCGCGCAACGAGACCGACCGCATCATCGGCTGGAGATGGGC 448
QY 389 GCAGATGACTACATCAACCAACCGTTCAACCGCGGTGAACCTGACGATTCGTGACGCAAC 448
DB 447 GCGACGACTACTGACCAAGCCCTTCTCCCGCGCAACTGCTTGGCCGCGATCAACGCC 388
QY 449 CTACTGTCCTCCATCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 508
DB 387 GTGCTGCGCGTACCGCGGATGCTGCGCGCAACCTCACCGTGAGCGAGCAGCGCGCTG 328
QY 509 TACAAGTTCAATGTTGGAACTGACATCAACAGCGCTTCTGTTGATCGGCTGATGCGG 568
DB 327 ATCGGCTTGGCGGAGTGGCACTGACACACGAGCGCGCCACCTGCTGACGACGCGCGC 268
QY 569 GAGCAGTACAAGTGGCGCGCAGCGAGTTCGCGCGCACTGCTTCACTTCTGTGAAACCCA 628
DB 267 ACCGTGTCGCGCTCAGCGCGCGAGTACCGCTGCTGCGGCTGCTTCTGATCATCG 208
QY 629 GCGAAATTCAGTCCCGTCTGAACTGCTGAAAGAAATGACCGCGCGTGAAGTGAACCG 688
DB 207 CAACGGGTACTCAGCGCGGACAGTTGCTCAACCTGACCCAGCGCGCGAGCGGACATC 148
QY 689 CACGACGCTACTGATGAGTGAAGTACGATCGCGCTGATTCGTAACATTTTGAATCTAGCGG 748
DB 147 TTCGACCGTTTCCATCGACCTGCTGTCAGCGCGCTGCGCCCAACGCTTGGCGACGCGC 88
QY 749 GATACGCGGAATCATCGCCACCATTCACGCTGAAGTTCGCTTCT 797
DB 87 CGCGAACCGGATCATCAAGACCGTGGCAGCGGCGCTATGTTCT 39

RESULT 10

US-09-252-991A-11722
; Sequence 11722, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11722
; LENGTH: 2166
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11722

Query Match 13.0%; Score 120.6; DB 3; Length 2166;

Best Local Similarity 49.2%; Pred. No. 9.3e-31;
Matches 349; Conservative 0; Mismatches 354; Indels 6; Gaps 1;
QY 95 GCAACATCGACACCGCCGACATCTTATCGTTGAAGACGAGTGTGTAACACGCAACACG 154
Db |||||
QY 1423 GCAATGGAGCATGTGATCACATCTGATCGTGCAGATGACCGGAGATTCGCGAGCTG 1482
Db |||||
QY 155 TTGAAGATTTTTCGAGCGGAAGGCTATGATGTTTTTCGAAGACGACAGATGCGCGGAA 214
Db |||||
QY 1483 GTAGGCACTACCTGAAGAGAAACGGCCTGCGCACCATCGTGCAGATGCGCGGAG 1542
QY 215 ATGCATCAGATCTCTCTCAATATGACATCAACCTTGGTATCATGATATCAATCTGCGG 274
Db |||||
QY 1543 ATGCGCGGTTCTCGAGCCAAACGCGTGGACCTGATGCTCTCGACATCATGATGCC 1602
QY 275 GGTAGAAGCGGTTCTGTTAGCGGTGAACCTGCGG-----AGCAGCGAATGTTGG 328
Db |||||
QY 1603 GCGGACGAGCGGCTGCTGCTGCGCGAGCTGCGGCTGCGGCAAGCACAAGGCCACGCG 1662
QY 329 TTGATGTTCTGACTGGCGGTGACACGAAGTGCATAAATTTCTGGGCTCGAATCGGT 388
Db |||||
QY 1663 GTGCTGATGCTCACCGCGCGCAACGACGAGACCGCATCATCGGCTTGAGATGGC 1722
QY 389 GCAGATGACTACATACCAAAACGTTTCAACCGCGTGAACCTGACGATTGTCGACGCAAC 448
Db |||||
QY 1723 GCGGAGACTACTGACCAAGCCCTTCTCGCCCGGAACTGCTTGGCCGATCAACGCC 1782
QY 449 CTACTGTCCCGTACCATGAATCTGGGTACTGTGACGGAAGACGTGTAGCGTTGAAAGC 508
Db |||||
QY 1783 GTGCTGCGGCTACGCGGATGCTGCGCGCAACCTCACCGTGAGGAGAGCGGCTG 1842
QY 509 TACAAGTTCAATGGTTGGAACTGACATCAACAGCGGTTGTTGATGGCCCTGATGGC 568
Db |||||
QY 1843 ATCGGCTTCCGCGAGTGGCACTGGACACGAGCGCCGCACTGCTCGACGACCGCGC 1902
QY 569 GAGCAGTACAAGCTGCGCGGACGAGTTTCCGCGCATCTTCACTTGTGAAAACCCA 628
Db |||||
QY 1903 ACGTGTGCTGCTCAGCGCGCGGAGTACCGCTGCTGCGGTTTCTCGATCATCG 1962
QY 629 GCGAAATTCAGTCCCGTCTGAACTGCTGGAAGAAATGACCGCGGTGAGTGAACCG 688
Db |||||
QY 1963 CAACGGGTACTCAGCGCGACCACTGCTCAACCTGACCGAGCGCGGAGCGGACATC 2022
QY 689 CAGGACGCTAGTACGCTGACGATCCGCGGTATTCGTAACATTTGCAATCTACGCG 748
Db |||||
QY 2023 TTGACCGTTTCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2082
QY 749 GATACCGCGGAAATCATCGCCACCATTCACGGTGAAGGTTATCGTTCT 797
Db |||||
QY 2083 CGGAAACCGGATACATCAAGACCGTGCAGGAGGGCTATGTGTTCT 2131

RESULT 11
US-09-252-991A-11865
; Sequence 11865, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11865
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11865

Query Match 12.3%; Score 113.8; DB 3; Length 2004;
Best Local Similarity 50.6%; Pred. No. 2.1e-28;
Matches 305; Conservative 0; Mismatches 292; Indels 6; Gaps 1;
QY 201 CAGATGGCGCGGAAATGATCAGATCCCTCTCTGAATATGATCAACCTGGTGATCATGG 260
Db |||||
QY 66 CCGATGGCGCGGAGATGCGCGGTTCTTCGAAGCAACACACCGTGGACCTGATCGTCTCG 125
Db |||||
QY 261 ATATCAATCTCCCGGTTAAGAACGCTTCTTGTAGCGCGTGAACCTGGCG-----AGC 314
Db |||||
QY 126 ACATCATGATGCCGCGGACGACGCGCTGCTCTGCGCGAGCTGCGCGTCCGCAAGC 185
QY 315 AGGCGAATGTTGCGTGTGATGTTCTGATGCGCGTGCACAAAGTGCATATAAATTTCTCG 374
Db |||||
QY 186 AAGAGGCCACCGCGGTGCTGATGCTCACCGCGGCAACGACGAGACCGCATCATCG 245
QY 375 GCCTCGAAATCGGTGACGATGACTACATCAACAAACCGTTCAACCGCGTGAACCTGACGA 434
Db |||||
QY 246 GCCTGAGATGGCGCGGACGACTACCTGACCAAGCCCTTCTCGCGCGGAACTGCTTG 305
QY 435 TTGTTGACGCAACCTACTGTCCTGCTACATGAATCTGGGTACTGTCAGCGAAGAACGTC 494
Db |||||
QY 306 CCGGATCAACGCGCTGCTGCGGCTGACGCGATGCTGCGCGCAACCTACCGTGAGCG 365
QY 495 GTAGGTTGAAGCTACAAGTTCAATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 554
Db |||||
QY 366 AGAGCAGCGCGCTGATCGGCTTTCGCGGAGTGGCAACGACACCGCGCCACCTGCG 425
QY 555 TCGGCGCTGATGGGAGCAGTACAAGCTGCGCGGAGGAGTTCGCGCGCATGCTTCACT 614
Db |||||
QY 426 TCGAGCAGCGCGGACCGTGGTTCGCGCTCAGCGGCGCGAGTACGCGCTGCGGCTGT 485
QY 615 TCTGTGAAAAACCGAGGCAAAATTCAGTCCCGTGTGAACTGCTGGAAGAAATGACCGGCC 674
Db |||||
QY 486 TCTCGATCATCCGCAACGGGTACTCAGCGCGACAGTGTCTCAACCTGACCGAGGCGC 545
QY 675 GTAGCTGAACCCGACGACCGTACTGTAGACGTGACGATCGCGGCTATTCGTAACATT 734
Db |||||
QY 546 GCGAGCGGAGCATCTTCGACCGGTTCCATCGACCTGCTGTCAGCGCGCTGCGCCCAACGCC 605
QY 735 TCGAATCTACCGCGGATACGCGGAAATCATCGCCACCATTCACCGTGAAGGTTATCGCT 794
Db |||||
QY 606 TCGGCGACGCGCGCGAACCGGAGTACATCAAGACCGTGGAGCGAGGCGCTATGTGT 665
QY 795 TCT 797
Db |||||
QY 666 TCT 668

RESULT 12
US-09-418-980-1
; Sequence 1, Application US/09418980
; Patent No. 6713071
; GENERAL INFORMATION:
; APPLICANT: Campos, Manuel
; APPLICANT: Baarsch, Mary Jo
; APPLICANT: Rosey, Everett
; APPLICANT: Ankenbauer, Robert
; APPLICANT: Warren-Stewart, Lynn
; APPLICANT: Suiter, Brian
; APPLICANT: Keach, Robin
; TITLE OF INVENTION: NOVEL PROTEINS FROM ACTINOBACILLUS PLEUROPNEUMONIAE
; FILE REFERENCE: PC9854A
; CURRENT APPLICATION NUMBER: US/09/418,980
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Actinobacillus pleuropneumoniae
; FEATURE:

; NAME/KEY: CDS
; LOCATION: (272)..(787)
US-09-418-980-1

Query Match 11.6%; Score 107.4; DB 3; Length 1018;
Best Local Similarity 80.3%; Pred. No. 2.3e-26;
Matches 126; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 662 AAAATGACCGCGGTGAGCTGAACCGCACGACCGTACTGTAGAGCTGACGATCGCGGT 721
Db 68 AAAATGACCGCGGTGAGCTGAACCGCACGACCGTACTGTGATGACGATCGCGGT 127

Qy 722 ATTGCTAAACATTTGCAATCTACGCGCGGATAGCGCGGAAATCATCGCACCATTCACGGT 781
Db 128 ATTGCTAAACATTTGCAATCTACGCGCGGATAGCGCGGAAATCATTCGTAACCATTCATGGT 187

Qy 782 GAAGGTTATCGTTCTGCGGTGATCTGGAAGTTAAT 818
Db 188 GAAGGTTATCGTTCTGCGGTGATCTGGAAGTTAAT 224

RESULT 13
US-09-634-238-133/C
; Sequence 133, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna C.
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.
; FILE REFERENCE: 11000.1043U1
; CURRENT APPLICATION NUMBER: US/09/634,238
; CURRENT FILING DATE: 2006-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-133

Query Match 10.6%; Score 98.6; DB 3; Length 729;
Best Local Similarity 46.3%; Pred. No. 2.1e-23;
Matches 323; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

Qy 156 TGAAGATATTTTGAAGCGGAGCGTATGATGTTTTCGAAGCGACAGATGGCGGAA 215
Db 728 TCAAGTTAAATTTGGATAAAGAGGCTATGCTGCTCATCATATGATGCGGAAG 669

Qy 216 TGCATCAGATCTCTCTGAATATGACATCAACCTGGTATCATGATATCAATCTGCCG 275
Db 668 CTTGAAGAAAGTTGAATCAGAAATCACCTGATTTGATTTTCTGATTTGATTTACCGA 609

Qy 276 GTAAGAACGGTCTTCTGTTAGCGGTGAATCTGCGGAGCAGCGGAATGTTGCGTTGATGT 335
Db 608 AAATTTGATGGGCTTGAAGTTGCCCGACAAATTCGTAAGACATGACACCGCATTTA 549

Qy 336 TCTGATGCGCGGTGACACGAAGTTCGATAAAATTTCTCGGCTCGAAATCGGTGAGATG 395
Db 548 TGTGACGCGGAAGGATTCGAAATTTGATAAAGTGTAGGACTGAGCTTGGGGCAGATG 489

Qy 396 ACTACATCACCAACCGTTCAACCGCGGTGACATGACGATTCGTGCAACCACTACTGT 455
Db 488 ATTATGTATCAACGCGGTTTTCCTCAATCGAAGACTTTGTCGCGGCGTGAAGGCTAATCTGC 429

Qy 456 CCGTACCATGAATCTCGGTACTGTCTAGCGGAAGAACGTCGTAGCTTTGAAAGCTACAGT 515
Db 428 GCCGTACTCATCAGCCAATGCTGCGGCACTGAGGAAGACGAAGCCAATCGTGAATTGG 369

Qy 516 TCATGTTGGGAAGTGGACATCAACAGCCGTTGTTGATCGGCCCTGATGGCGAGCT 575
Db 368 AAGTTGGCGAATTTAACGATTCATCCGACGCGTATACGGTATCTAAACGCGGTGAGAATA 309

Qy 576 ACAAGCTCGCGCGCAGCGAGTTCGCGCCCATGCTTCACTTCTGTGAAACCCAGGCAAAA 635
Db 308 TCGAATTAACCCACCGGAGTTTGAATTTGTCGATTACTTGTCCCGCCACTTGGGACAGG 249

Qy 636 TTCAAGTCCGCTGCTGCTGAAAGAAATGACCGCGCGTGAAGTGAACCCGACGACC 695
Db 248 TTATGACCGCGGAACATCTGCTACAGACAGTTTGGGGCTATGACTACTTTGGCGATGTCC 189

Qy 696 GTACTGTAGAGTACGATCGCGGTATTTGTAACATTTTCAATCTACGCGCGGATGCG 755
Db 188 GGACTGTGGATGTGACGGTTGCGCCCTCGCGGAGAAATCGAAGACAATCCGTCCTCCATC 129

Qy 756 CGGAATCATCGCCACCAATTCACGGTGAAGTTTATCGCTTCTGCGGTGATCTGGAAGATT 815
Db 128 CTGAGTGTGTTGTCACCGGCGCGGTAGGCTATTATTTCGGAATCTTCAACGCTGAAT 69

Qy 816 AATCGGCTTTACCAACCGTCAAAAAAAGCGCGCTTTT 852
Db 68 AATTAGGAGCACCACTGTTGAATAAAAAAGATTGCTT 32

RESULT 14
US-09-252-991A-6307
; Sequence 6307, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6307
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6307

Query Match 10.5%; Score 97.2; DB 3; Length 762;
Best Local Similarity 49.4%; Pred. No. 6.7e-23;
Matches 339; Conservative 0; Mismatches 338; Indels 9; Gaps 3;

Qy 114 ACATTCTTATCGTTGAAGACGAGTTGGTAAACACGCAACACCGTTGAAAAGTATTTTGAAG 173
Db 53 AGATCCTCGTGTGACGACGACGCGCCGCTGCGCGCTCTCGAACGTTTCTCTCGACG 112

Qy 174 CGGAAGCTATGATGTTTTTGAAGCGACAGATGGCGGGAATGATGATCAGATCTCTCTG 233
Db 113 AGCAGGGCTACCGGCTCGCGCGGTGGAGAACACCGAGCAGATGATCGCTTCTATCCC 172

Qy 234 AATATGACATCAACCTGTGATCATGATATCAATTCGCGGGTGAAGAACGCTTCTGT 293
Db 173 GCGAATCTGTTCCATGTTGGTTGCTGCTCGACCTGATGCTTCCCGGAGACGCGCTTCACTG 232

Qy 294 TAGCGGTGAACCTCGCGGAGCAGCGCAAT---GTTGCGTTGATGTTCTGACTGGCGGTG 350
Db 233 CTTGCGCGGCTGCGCGAACAACAATCAGGTGCGGATCATCATGCTCACCGCAAGG 292

Qy 351 ACAACGAAGTCGATAAAATTTCTCGGCTCGAAATCGGTGCGAGTACTACATCACCAAC 410

Db 293 GCGACGAGCGGACCGCATCCAGGCGCTGGAACTGGGCGCGACGACTACCTGGCCAAAGC 352
Qy 411 GCTTCAACCCGGTGAACGTGACGATTCGTGACGCAACCTACTGTCCGTACCATGAATC 470
Db 353 GCTTCAACCCGGCGAACTGCTGGCGCGGATCAAGGCGGTGCTGCGCGCCGACGAGCGCGC 412
Qy 471 TGGGTACTGTGACGCAAGAAAGCTGCTAGCGTTGAAAGCTACAAAGTTCAATGGTTGGGAAC 530
Db 413 TGGTCC---CGGCGCGCTGCGGTGCGGACGAGGTGGTACCTTCGGGACTACCGAC 469
Qy 531 TGGACATCAACAGCGCTTGGTTGATCGGCGCTGATCGGCGAGCAGTACAAAGCTCGCGCGCA 590
Db 470 TGTTCCTCGCCACCGCGAACTGAAG---AAGGGCGACGAGTACACATGCTGACCAACG 526
Qy 591 GCGAGTTCCGCCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTGTG 650
Db 527 GCGAGTTCCCGTGTCAAGGCGCTGGTGCAGACGCGCGCGAAACCGTGCACCGCGGACA 586
Qy 651 AACTGCTGAAGAAATGACCGCGGTGAGCTGAAACCGCACGCGTACTGTAGACGTGA 710
Db 587 AGCTGATGAACCTCGCGCGTGGCGGAGTGGGATGCCCTGGAGGTTCCATCGACGTGC 646
Qy 711 CGATCCGCGGTATTCGTAAACATTTGCAATCTACGCGGATACGCGCGAAATCATCGCCA 770
Db 647 AGATCTCGCGCTGCGCGGTGATCGAGCGGATCCGTCCAAAGCGCGCTATATCCAGA 706
Qy 771 CCAITTCAGGTGAAGGTTATCGCTTC 796
Db 707 CGGTCTGGGGCGTGGCTACGTGTTTC 732

RESULT 15

US-09-134-001C-226
; Sequence 226, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 226
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-226

Query Match 10.4%; Score 96.2; DB 3; Length 705;
Best Local Similarity 48.5%; Pred. No. 1.4e-22;
Matches 330; Conservative 0; Mismatches 343; Indels 8; Gaps 2;

Qy 117 TTCTTATCGTTGAAGACGAGTTGGTAAACGCAACACGTTGAAAGTATTTTCGAAGCGG 176
Db 17 TTGTTGTAGTTGACGATGAAACCAATTTGCTGATATTTAGAAATTTAAATTTAAAAAAG 76
Qy 177 AAGGCTATGATGTTTTCGAAGCGACAGATGGCGGGAATGCATCAGATCCTCTCTGAAT 236
Db 77 AAGTTACGACGTATATTCGGCTTATGCGGTAATGACGAGTAAATCTATGAAG 136
Qy 237 ATGACATCAACCTGGTGAATCATGATATCAATCTGCGGGTGAAGACGGTCTTCTGTTAG 296
Db 137 AAGAACCAGATATCGTCTTACTTGTATATCATGTTACCTGTTAGAGATGTTATGGAAGTAT 196
Qy 297 CCGGTGAACCTGCGGACGAGGCGAATGTTGCGTTGATGTTCTGACTGCGCGTGACAAG 356
Db 197 GTCGTGAAGTGGTAAAAAGTAAATGATCCGATTTATCATGCTGACAGCAAGAGTTCTG 256

Qy 357 AAGTCGATAAAATCTCGGCTCCGAATCGGTGCAGATGACTACATACCAAAACCGTTCA 416
Db 257 AAATTGATAAAGTATTAGGCTTGAATTTAGGTGCAGATGATTAGTAACCAATTTA 316
Qy 417 ACCGCGTGAACCTGACGATTCGTGCACGCAACCTACTGTCCCGTACCAATGAATCTGGGTA 476
Db 317 GTACTCGTGAACCTCATCGCACGTG-----TGAAGCGAACTTACGCCGTCATTATTCA 369
Qy 477 CTGTCAAGGAAGAACGTCGTAGCGTTGAAAGCTACAAAGTTCAATGGTTGGGAACCTGGACA 536
Db 370 CAACGAGCTCAAGAAAGTAAAGTGGTGCACAAATGAATTTACAAATTAAGATATTGTGATT 429
Qy 537 TCAACAGCGG--TTGCTTGTATCGGCCCTGATGGCGAGCAGTACAAAGCTCCCGCGACGCGAG 595
Db 430 TATCCAGATGCATATTCAATTTAAAAAAGCTGGAAGACATTTGAATTAACGATCGTGAA 489
Qy 596 TTCCGCGCCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCCGTGGTGAACCTG 655
Db 490 TTCGAGCTATTTCATTTCTTCTAAACATATGGGTCAAGTCATGACACGCTGAACACTTA 549
Qy 656 CTGAAGAAATGACCGGCCGTCGAGCTGAAACCGCACGACCGTACTGTAGAGTGAAGATC 715
Db 550 CTACAAACAGTGTGGGGTTACGATTATTTTGGTGTATGTTCTGCTGTCGACGTAAACAATT 609
Qy 716 GCGGTATTTCGTAAACATTTTCGAATCTACGCGGATACGCGGAAATCATCGCCACCAATT 775
Db 610 GCGGTTTAAAGAAAGAAATTTGAAGATGATCCATCTCATCCAGAAATACATTTGTGACAGT 669
Qy 776 CACGCTGAAGGTTATCGCTTC 796
Db 670 AGAGCGTTGGATACTTCTC 690

Search completed: February 13, 2006, 14:41:47
Job time : 219 secs

C	1	927	100.0	927	8	US-10-613-990A-31	Sequence 31, Appl
	2	925.8	99.9	1732	6	US-10-264-237-983	Sequence 983, App
	3	508.6	54.9	759	8	US-10-613-990A-19	Sequence 19, Appl
	4	447.4	48.3	1048	9	US-10-450-763-28743	Sequence 28743, A
	5	320.6	34.6	191996	9	US-10-795-159-683	Sequence 683, App
C	6	319.4	34.5	1830121	7	US-10-329-670-1	Sequence 1, Appl
	7	319.4	34.5	1830121	8	US-10-158-865-1	Sequence 1, Appl
C	8	319.4	34.5	1830121	9	US-10-981-687-1	Sequence 1, Appl
	9	316.2	34.1	711	3	US-09-815-242-7046	Sequence 7046, Ap
	10	332.8	34.1	711	7	US-10-283-132A-22157	Sequence 22157, A
	11	216.2	25.1	1110	9	US-10-450-763-18226	Sequence 18226, A
	12	161.6	17.4	2125	9	US-10-795-159-291	Sequence 291, App
C	13	153.6	16.6	2227	9	US-10-795-159-230	Sequence 230, App
	14	107.4	11.6	1018	8	US-10-770-834-1	Sequence 1, Appl
C	15	106.8	11.5	11275	6	US-10-240-689-39	Sequence 39, Appl
	16	103.8	11.2	714	7	US-10-283-132A-35444	Sequence 35444, A
	17	101.4	10.9	705	7	US-10-283-132A-35988	Sequence 35988, A
	18	99	10.7	732	7	US-10-283-132A-14497	Sequence 14497, A
	19	98	10.6	714	7	US-10-283-132A-24570	Sequence 24570, A
C	20	96.8	10.4	2256646	7	US-10-470-565-1	Sequence 1, Appl
	21	96.2	10.4	705	7	US-10-724-972A-3136	Sequence 3136, Ap
	22	96.2	10.4	721	7	US-10-283-132A-34637	Sequence 34637, A
	23	95	10.2	1163020	7	US-10-398-221-10	Sequence 10, Appl


```
QY 301 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGACTGSCCGTGACAACGAAGT 360
Db 301 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGACTGSCCGTGACAACGAAGT 360
QY 361 CGATATAAATTTCTCGGCTCGAATCGGTGACAGTACATCACCACAAACGGTTCAACCC 420
Db 361 CGATATAAATTTCTCGGCTCGAATCGGTGACAGTACATCACCACAAACGGTTCAACCC 420
QY 421 GCGTGAACCTGACGATTCGTGACGCAACCTACTGTCCTGACCTACCATGAATCTGGGTACTGT 480
Db 421 GCGTGAACCTGACGATTCGTGACGCAACCTACTGTCCTGACCTACCATGAATCTGGGTACTGT 480
QY 481 CAGCGAAGACCTGCTAGCGTTGAAGCTTACAGTTCAATGTTGGGAACTGGACATCAA 540
Db 481 CAGCGAAGACCTGCTAGCGTTGAAGCTTACAGTTCAATGTTGGGAACTGGACATCAA 540
QY 541 CAGCGGTTGTTGATCGGCGCTGATGGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCCG 600
Db 541 CAGCGGTTGTTGATCGGCGCTGATGGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCCG 600
QY 601 CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGCTGAA 660
Db 601 CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGCTGAA 660
QY 661 GAAATGACCGCGCTGAGCTGAAACCGCAGCAGTACGCTGCTAGACGTGACGATCCGCG 720
Db 661 GAAATGACCGCGCTGAGCTGAAACCGCAGCAGTACGCTGCTAGACGTGACGATCCGCG 720
QY 721 TATTGCTAAACATTTGGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 780
Db 721 TATTGCTAAACATTTGGAATCTACGCGGATACGCGGAAATCATCGCCACCATTCACGG 780
QY 781 TGAAGGTTATCGCTTCTCGGCTGATCGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 840
Db 781 TGAAGGTTATCGCTTCTCGGCTGATCGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 840
QY 841 AACGGGCTTTTAGCGCGTTTATTTTCAACCTTATTTCCAGATACGTAACATCATC 900
Db 841 AACGGGCTTTTAGCGCGTTTATTTTCAACCTTATTTCCAGATACGTAACATCATC 900
QY 901 GTCCGTTGTAACCTCTTTACTGGCTTT 927
Db 901 GTCCGTTGTAACCTCTTTACTGGCTTT 927
```

RESULT 2

```
US-10-264-237-983/c
; Sequence 983, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 983
; LENGTH: 1732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1568)..(1631)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1631)..(1631)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1673)..(1673)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1701)..(1701)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-264-237-983
```

```
Query Match 99.9%; Score 925.8; DB 6; Length 1732;
Best Local Similarity 99.7%; Pred. No. 7.4e-300;
Matches 924; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATGTTACCCCGATCATGTTAATTTGACAGATGCAATCAGGAGGTTCAGGACTTTTGT 60
Db 1313 GTCATGTTACCCCGATCATGTTAATTTGACAGATGCAATCAGGAGGTTCAGGACTTTTGT 1254
QY 61 ACTTCCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAATGCAAGCCCGCACATTC 120
Db 1253 ACTTCCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAATGCAAGCCCGCACATTC 1194
QY 121 TATCGTTGAAGACGAGTTGGTAAACGCAACCGTTGAAAAAGTATTTTCGAAGCGGAAG 180
Db 1193 TATCGTTGAAGACGAGTTGGTAAACGCAACCGTTGAAAAAGTATTTTCGAAGCGGAAG 1134
QY 181 CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCAATCAGATTCCTCTGAATATGA 240
Db 1133 CTATGATGTTTTCGAAGCGACAGATGGCGCGGAAATGCAATCAGATTCCTCTGAATATGA 1074
QY 241 CATCAACCTGCTGATCATGGATATCAATCTCGCGGTTAAGAACGGTCTTCTGTTAGCGCG 300
Db 1073 CATCAACCTGCTGATCATGGATATCAATCTCGCGGTTAAGAACGGTCTTCTGTTAGCGCG 1014
QY 301 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGACTGSCCGTGACAACGAAGT 360
Db 1013 TGAACCTGCGGAGCAGCGAATGTTGCGTTGATGTTCTGACTGSCCGTGACAACGAAGT 954
QY 361 CGATAAATTTCTCGGCTCGAATCGGTGACAGTACATCACCACAAACGGTTCAACCC 420
Db 953 CGATAAATTTCTCGGCTCGAATCGGTGACAGTACATCACCACAAACGGTTCAACCC 894
QY 421 GCGTGAACCTGACGATTCGTGACGCAACCTACTGTCCTGCTACCATGAATCTGGGTACTGT 480
Db 893 GCGTGAACCTGACGATTCGTGACGCAACCTACTGTCCTGCTACCATGAATCTGGGTACTGT 834
QY 481 CAGCGAAGACCTGCTAGCTGTTGAAGCTACAAGTTCAATGTTGGGAACTGGACATCAA 540
Db 833 CAGCGAAGACCTGCTAGCTGTTGAAGCTACAAGTTCAATGTTGGGAACTGGACATCAA 774
QY 541 CAGCGGTTGTTGATCGGCGCTGATGGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCCG 600
Db 773 CAGCGGTTGTTGATCGGCGCTGATGGGAGCAGTACAAGCTGCGCGCGCAGCGAGTTCCG 714
QY 601 CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGCTGAA 660
Db 713 CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGCTGAA 654
QY 661 GAAAATGACCGCGCTGAGCTGAAACCGCAGCAGTACGCTGCTAGACGTGACGATCCGCG 720
Db 653 GAAAATGACCGCGCTGAGCTGAAACCGCAGCAGTACGCTGCTAGACGTGACGATCCGCG 594
QY 721 TATTGCTAAACATTTGGAATCTTACGCCGGAATCATCGCCACCATTCACCG 780
Db 593 TATTGCTAAACATTTGGAATCTTACGCCGGAATCATCGCCACCATTCACCG 534
QY 781 TGAAGGTTATCGCTTCTCGGCTGATCGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 840
Db 533 TGAAGGTTATCGCTTCTCGGCTGATCGGAAGATTAAATCGGCTTTACCAACCGTCAAAAA 474
QY 841 AACGGGCTTTTAGCGCGCTTTTATTTTCAACCTTATTTCCAGATACGTAACATCATC 900
```

Db 473 AACGGCGCTTTTAGCGCGTTTTTATTTTCAACCTTATTTCCAGATACGTAACTCATC 414
Qy 901 GTCGGTTGTAACCTCTTTACTGGCTTT 927
Db 413 GTCCGKTGTAACTCTTTTACTGGCTTT 387

RESULT 3
US-10-613-990A-19
; Sequence 19, Application US/10613990A
; Publication No. US20040180404A1
; GENERAL INFORMATION:
; APPLICANT: Ajinomoto Co., Inc.
; TITLE OF INVENTION: Method for Producing Target Substance by Fermentation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/613,990A
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: JP 2002-203764
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 759
; TYPE: DNA
; ORGANISM: Pantoea ananatis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (41)..(757)
US-10-613-990A-19

Query Match 54.9%; Score 508.6; DB 8; Length 759;
Best Local Similarity 80.6%; Pred. No. 1.2e-159;
Matches 607; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

Qy 69 TTTCGATTAGTTGGC--AATTTAGTAGCAAAAGATGAGACCCCGGCACTTCTATCGT 126
Db 7 TTTCGATTAGTTGGCAAAATTTAGTAGCTAAACATGACAGCCCGGCACTTCTATCGT 66

Qy 127 TGAACAGAGTTGGTAAACGCAACAGCTGTGAAAAGTATTTTCGAAGCGGAAGGTATGA 186
Db 67 TGAACAGAGTTGGTCAAGGATATCCCTCNAAGCAATTTTGAAGCGGAAGGTATGT 126

Qy 187 TGTTTTCGAAGCAGATGCGCGGAAATGCATCAGATCTCTCTGTAATATGACATCAA 246
Db 127 CGTGACGAAGCGAGTGTGACAGATGACACGAGTGTGACGCAATGATGTCAA 186

Qy 247 CTGTGATCATGATATCAATCTCGCGGTGAAGAACGCTCTTCTGTTAGCGCGTGAAT 306
Db 187 TCTGGTTATATGAGATCAATCTCGCGGTGAAGAACGCGCTGTTACTGGCAGCGTGAAT 246

Qy 307 GCGGAGCAGGCGAATGTTGGCTTGTGATGTTCTGACTGCGGTGACACGAGTGCATAA 366
Db 247 GCGTGAAGCAATGTCGATGTTGATGTTCTGACCGGACGCAATAGCAATGTCATAA 306

Qy 367 AATTCCTCGGCTCGAATCGGTGACGATGACTACATCACCAACCGTTCAACCCCGTGA 426
Db 307 AATTCCTGGGCTGGAAATTTGGTGACAGGACTACATTAAGCCGTTTAAACCCAGCGA 366

Qy 427 ACTGACGATCTGTGACGCAACCTACTGTCGGTACCATGATCTGGGTACTGTGACGGA 486
Db 367 ATTAACATATCTGTGACGCTAACCTGCTGTGGCCACCATGATTTGCGCTTACCCCAATGA 426

Qy 487 AGAAGCTCTAGCTGTGAAGCTACAGTTCAATGTTGGGAACCTGGGACATCAACAGCGG 546
Db 427 AGAGCGTCCGCAAGTTGAAGCTATAAGTTCAACCGCTGGGAGCTGGACATCAACAGCGG 486

Qy 547 TTCTGTTGATCGGCTCTGATGGCGAGCAGTACAGCTGCGCGCAGCGAGTTCCCGGCCAT 606
Db 487 CTCACCTCATCAATCCCAAGCGTGAAGCTACAACTGCGCGCAGTGAAGTTCCCGTCCAT 546

Qy 607 GCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCCGTGTGCTGTAAGTAAAT 666
Db 547 GCTGCACTTCTGCGAAAATCCCGGCAAGATTCAGACGCGGTGCTGATTTGCTGAAGAAAT 606

Qy 667 GACCGCGGTGAGCTGAAAACGACGACCGTACTGTAGACGTGACGATCCGCGGTATTCG 726
Db 607 GACCGGAGCGGATCTCAAGCAACGACGCGTACTGTTGACGTGACCAATCCGTCGTATCG 666

Qy 727 TAAACATTTGAAATCTACGCGGATAGCCGGAATCATCGCCACCATTCACGGTGAAGG 786
Db 667 TAAACATTTGAAATCTACGCGGATAGCCGCTGAATCATCGCCACCATTCACGGGGAAGG 726

Qy 787 TTATCGCTTCTGCGGTGATCTGGAAGATTAATC 819
Db 727 TTATCGTTTCTGTGTGACCTGCGAGGATTAAGC 759

RESULT 4
US-10-450-763-28743
; Sequence 28743, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 28743
; LENGTH: 1048
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (569)..(1021)
; OTHER INFORMATION: 98% homologous to Escherichia coli dye, accession number
; OTHER INFORMATION: M10044, Smith-Waterman Score=775.
US-10-450-763-28743

Query Match 48.3%; Score 447.4; DB 9; Length 1048;
Best Local Similarity 99.8%; Pred. No. 5.5e-139;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 312 AGCAGGCGAATGTTGCGTTGATGTTCTTGACTGGCCGTGACACGAAAGTCGATAAAATTC 371
Db 573 AACAGGCGAATGTTGCGTTGATGTTCTTGACTGGCCGTGACAAACGAAAGTCGATAAAATTC 632

Qy 372 TCGGCTCGAATCGGTGCAGATGACTACATCACCACCGTTCAACCCGCGTGAACCTGA 431
Db 633 TCGGCTCGAATCGGTGCAGATGACTACATCACCACCGTTCAACCCGCGTGAACCTGA 692

Qy 432 CGATTCTGTGACGCAACCTACTGTCCCGTACATGAATCTGGGTACTGTGACGCGAAGAAC 491
Db 693 CGATTCTGTGACGCAACCTACTGTCCCGTACATGAATCTGGGTACTGTGACGCGAAGAAC 752

Qy 492 GTGCTAGCGTTGAAAGCTTACAAAGTTCAATGTTGGGAACCTGGAACATCAACAGCCGTTGTT 551
Db 753 GTGCTAGCGTTGAAAGCTTACAAAGTTCAATGTTGGGAACCTGGAACATCAACAGCCGTTGTT 812

Qy 552 TGATCGGCGCTGATGGCGGAGCAGTACAGCTGCGCGCAGCGAGTTCCGCGCATGCTTC 611
Db 813 TGATCGGCGCTGATGGCGGAGCAGTACAGCTGCGCGCAGCGAGTTCCGCGCATGCTTC 872

Qy 612 ACTTCTGGAACAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAAGAAAATGACCG 671
Db 873 ACTTCTGGAACAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAAGAAAATGACCG 932

Qy 672 GCGCTGAGCTGAAACCCGACGACCGATCTGTAGAGTGAAGTCCGCGGTATTCGTTAAAC 731

Db 933 GCGTGAGCTGAAACCGCAGCAGCGTACTGTAGAGTGACGATCGCGCTATTTCGTAAAC 992

QY 732 ATTTCGAATCTACGCGCGGATACGCGGAA 760

Db 993 ATTTCGAATCTACGCGCGGATACGCGGAA 1021

RESULT 5

US-10-795-159-693/c

; Sequence 683, Application US/10795159

; Publication No. US20050221439A1

; GENERAL INFORMATION:

; APPLICANT: BAKALETZ et al.

; TITLE OF INVENTION: GENES OF AN OTITIS MEDIA ISOLATE OF NONTYPEABLE H. INFLUENZAE

; FILE REFERENCE: 28335/38815A

; CURRENT APPLICATION NUMBER: US/10/795,159

; PRIOR FILING DATE: 2004-03-05

; PRIOR APPLICATION NUMBER: US 60/453,134

; PRIOR FILING DATE: 2003-03-06

; NUMBER OF SEQ ID NOS: 771

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 683

; LENGTH: 191996

; TYPE: DNA

; ORGANISM: H. influenzae

US-10-795-159-693

Query Match 34.6%; Score 320.6; DB 9; Length 191996;

Best Local Similarity 66.0%; Pred. No. 4.9e-95;

Matches 480; Conservative 0; Mismatches 244; Indels 3; Gaps 1;

QY 85 AATTAGGTAGCAACATCGACACCGCCGACATTTCTATCGTTGAAGACGAGTTCGTAAC 144

Db 55869 AGTATAAGGATGGAAATGACTCTCCAAAATTCGTGTGTGAAGATGAATTTGTCCAC 55810

QY 145 ACGCAACACGTTGAAAGTAATTTTGAAGCGGAAGCTATGATGTTTTCGAAGCGACAGA 204

Db 55809 TCGAATACGCTTAAGGATTTTGNAGCGGAGGATGATGTTTGAAGCAGAAA 55750

QY 205 TGGCGGGAATGCATCAGATCCTCTCTGAAATATGACATCAACCTGGTGATCATGGATAT 264

Db 55749 TGGTGTGAAATGCAATCATATATGCAATATGCAATATAATTAATTTGGTTGTGATGGATAT 55690

QY 265 CAATCTGCGGTAGAACGGTCTCTGTAGCGGTGAACCTGCGGACGAGCGCAATGT 324

Db 55689 TAATTTACCGGAAACCGGCTTATTTGCAAGAGAACTCCGTAAGAAATTAAGCTT 55630

QY 325 TCGCTTGATGTTTCTGACTGGCGGTGACAAACGAAGTCGATAAAATTTCTGGCCTCGAAAT 384

Db 55629 ACCTCTTATTTTAACTGGTGGAGATTAATGAATGGATATAAATTTTAGGTCGGAAT 55570

QY 385 CGGTGCAGATGATCATACACCAACCGTTCAACCGCGTGAACTGACGATTCGTGCGACG 444

Db 55569 TGGTCCGACGATTAATTTAAACCAACCTTTAAACCTAGAGAACTGACTATCCGCGCAGC 55510

QY 445 CAACCTACTGTCGCTGACCATGAACTCGGTACTGTTCAGCGAAGAACGTCGTAGCGTTGA 504

Db 55509 TAATCTATTGCTGCGACAATG---TCGCATCAAGAAAAGAAATATACATTTGGTCGAGA 55453

QY 505 AAGCTACAAGTTCAATGGTTGGGAACTGGACATCAACAGCCGTTGTTGATCGGCCCTGA 564

Db 55452 ATTCTATCGTTTAACTGGTTGGAAATTAGCCCTAAATAGCCACAGTTTAATTAACACAGA 55393

QY 565 TGGCGAGAGTACAGCTGCGCGGACGAGGTTCCGCGCCATGCTTCACCTCTCTGTGAANA 624

Db 55392 AGGACAAGAAATTCAACTTCTCGCAGTGAATTTCTGTCGAATGTTACATTTCTGTGAANA 55333

QY 625 CCAGGCAAAATTCAGTCCCGTCTGAACTGCTGAGAGAAATGACCGCGGTGAGCTGAA 684

Db 55332 TCCAGGAAATTTGCAACCGCGAGAAATTTGCTGAAAAAATGACGGGACGAGAGTTAAA 55273

QY 685 ACCGCAACGACCGTACTGTGATGACGATCCGCGGTATTCGTAAACATTTTCGAATCTAC 744

Db 55272 ACCTCAGGATCGTACCCTAGATGTCAAAATTCGACGTATCAGAAAACATTTTGAAGATCA 55213

QY 745 GCGGATACGCGGAAATCATCGCCACCATTCACGGTGAAGTTATCGCTTCTCGCGTGA 804

Db 55212 TCCAAATACTCCAAATATATCATTTAGCTATACATGGAGAGGCTATCGCTTTTGGGAGA 55153

QY 805 TCTGAA 811

Db 55152 TATTGAA 55146

RESULT 6

US-10-329-670-1/c

; Sequence 1, Application US/10329670

; Publication No. US20040018503A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag

; TITLE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: PB186P1

; CURRENT APPLICATION NUMBER: US/10/329,670

; CURRENT FILING DATE: 2002-12-24

; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: US 08/487,429

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1830121

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4747)..(4747)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (9921)..(9921)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (10150)..(10150)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (29298)..(29298)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (36543)..(36543)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (36551)..(36551)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (36636)..(36636)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (40808)..(40810)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (44416)..(44416)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (44905)..(44905)

```
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)
```

```
Query Match 34.5%; Score 319.4; DB 7; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4,4e-94;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
QY 85 AATTTAGGTAGCAAAACATGCAGACCCCGCACATCTTTATCGTTGAAGACGAGTTGCTAAC 144
```

Db 936640 AGTATAAGGATGGAAATGACTACTCCAAAAATTCGTGTTGTAAGATGAATTTGTCAC 936581

Qy 145 ACGCAACACAGTTTGAATAAGTATTTTCGAAGCGGAGGCTATGATGTTTTTCGAAGCGACAGA 204

Db 936580 TCGAATACGCTTAAAGGGATTTTGAAGCGGAGGATATGATGTGTTTGAACGACGAAAA 936521

Qy 205 TGGCGCGGAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGATAT 264

Db 936520 TGGTGTGGAATGCATCATATATTTGGCAAAATCAATAATTAATTTGGTTGTGATGATAT 936461

Qy 265 CAATCTCCGGGTAAAGAACGGTCTTCTGTAGCGCGTGAACCTCGCGAGCAGCGCAATGT 324

Db 936460 TAATTTACAGGCAAAAACGGCTTATTAATTTGGCAAGGAACTCGGTGAAGAAATTAAGCTT 936401

Qy 325 TGGGTTGATGTTCTCTGACTCGGCGGTGACACGAAGTGCATAAAAATTTCTCGGCTCGAAAT 384

Db 936400 ACCTCTTATTTTAACTGGTCGAGATAAATGAATGGATAAAAAATTTAGGTTTGGAAAT 936341

Qy 385 CGGTGCAGATGACTACATCAACCAACCGTTCAACCGCGTGAACCTGACGATTCGTGCACG 444

Db 936340 TGGTCCGATGACTATTTTGAATGACTTAAGCTTTTAACTTGAAGAACTGACTATCCGCGCAG 936281

Qy 445 CAACCTACTGTCGGTACCAATGTAATCTGGGTACTGTCTCAGCGAAGAACTGCTAGCGTTGA 504

Db 936280 TAATCTATTGCATCGTCAATG---CCGCATCAAGAAAAAATAATACATTTGTCGAGA 936224

Qy 505 AAGCTACAAGTTCAATGGTTGGGAACTGGACATCAACAGCGGTTGTTGATCGGCCCTGA 564

Db 936223 ATTCATATGTTTAAATGGTTGGAAATTAGACCTTAATAGCCACAGTTTAATTAACACAGA 936164

Qy 565 TGGCGACAGTACAAGCTGCGCGCGACGAGTTCGCGCCCATGCTTCACATTCCTGTGAAAA 624

Db 936163 AGGACAGAATTCAACTCTCTCGCAAGTATTCGTGCAATGTACATTTCTGTGAAAA 936104

Qy 625 CCCAGGCAAAATTCAGTCCGCGTGTGCTGAACTGCTGGAAGAAAAATGACGGCGCTGAGCTGAA 684

Db 936103 TCCAGGAAAAATTCGCAACGCGCGAAGAAATTCGTGAAAAAATGACCGACGAGAGTTAA 936044

Qy 685 ACCGACGACCGTACTCTAGACGTGACGATCCGCGCGTATTCGTAACATTTCCGAATCTAC 744

Db 936043 ACCTCAGATGCTACCGTAGATGTCACAAATTCGACGATTCAGAAAAATTTTGAAGATCA 935984

Qy 745 GCCGGATACCGCGAAATCATTCGCCACCATTCACGGTGAAGGTTATCGCTTCTCGCGTGA 804

Db 935983 TCCCAATACTCCAAATATCATTAATGACTATACATGAGAAAGGCTATCGTTTTTGTGGAGA 935924

Qy 805 TCTGGAGA 813

Db 935923 TATTGAATA 935915

RESULT 7

US-10-158-865-1/c

; Sequence 1, Application US/10158865

; Publication No. US20040203093A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag

; TITLE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: P8186P2C1D1

; CURRENT APPLICATION NUMBER: US/10/158,865

; CURRENT FILING DATE: 2002-06-03

; PRIOR APPLICATION NUMBER: US 09/557,884

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: US 08/476,102

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: Patent version 3.1

; SEQ ID NO 1

; LENGTH: 1830121

```
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      34.5%; Score 319.4; DB 8; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.4e-94;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

Qy      85 AATTTAGTAGCAACATGACGACCCGACATCTTATCTGTTGAGACGAGTTGGTAAC 144
Db      936640 AGTATAAGGATGGAAATGACTACTCCAAAAATTTCTCGTTGTTGAAGATGAAATTTGTAC 936581

Qy      145 ACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGCACAGA 204
Db      936580 TCGAATACGCTTAAGGATTTTGAAGCGGAGGATATGATGTTTGAAGCGAAGAAA 936521

Qy      205 TGGCGCGGAAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGATCATGGATAT 264
Db      936520 TGGTGTGTAATGCATCATATATTGGCAATATATAATTAATTTGTTGTGATGGATAT 936461

Qy      265 CAATCTCGCGGTAAAGACGGTCTTCTGTTAGCGGTGAACCTGCGGAGCGAGCGGAATGT 324
Db      936460 TAATTTACAGGCAAAAAACGGCTTATTATTGGCAAGAGAACTCCGTGAAGAAATTAAGCTT 936401

Qy      325 TCGGTTGATGTTCTGACTGCGCGGTGACACGAAGTTCGATAAAATTTCTCGGCTCGAAAT 384
Db      936400 ACCTCTTATTTTTTAAGTCTCGAGATATGAATGATGATTAATTTTAGGTTGGAAT 936341

Qy      385 CGGTGCAGATGACTACATCACAAACCGTTCAACCCGCGTGAACCTGACGATTCGTGACG 444
Db      936340 TGGTCCGATGACTATTGACTAAGCCCTTTTAAACCTAGAGAACTGACTATCCCGGACG 936281

Qy      445 CAACCTACTGTCGCGTACCATGATCTGGGTACTGTCTGACGCAAGAACCTCTAGGCTTGA 504
Db      936280 TAATCTATTGTCATCTGTGCAATG---CCGCATCAAGAAAAAGAAAAATACATTTGGTCGAGA 936224

Qy      505 AAGCTACAAGTTCAATGTTGGGAACCTGGACATCAACAGCCGTTCTGTTGATCGGCCCTGA 564
Db      936223 ATTCTATCGTTTTAATGTTGGAATTAGACTTAATAGCCACAGTTTAAATTACACCAGA 936164

Qy      565 TGGCGAGCAGTACAAGCTGCGCGCGAGCGAGTTCCGCGCCATGCTTCACTTCTTGTGAAAA 624
```

```
Db 936163 AGGACAAGATTCAACTTCCTCGAGTGAAATTCGTGCAATGTTACATTTCTGTGAAAA 936104
QY 625 CCCAGGCAAAATTCAGTCCGCTGCTGAACTGCTGAAGAAATGACCGCGCGTGAAGCTGAA 684
Db 936103 TCCAGGAAATTCGCAACGCGCGAAGAAATTCGTGAAAAAATGACCGGACGAGAGTTAAA 936044
QY 685 ACCGCACACCGTACTAGTAGAGTACGATCCGCGGTATTCGTAACATTTTCAATCTTAC 744
Db 936043 ACCTCAGGATCGTACCGTAGATGTCACAAATTCGAGGTATCAGAAACATTTTGAAGATCA 935984
QY 745 GCGGATACGCGGAAATCATCGCCACCATTACCGTGAAGGTATCGTCTCTCGCGTGA 804
Db 935983 TCCCAATACTCCAATATCAATTCATGCTATACATGATGAGAGAGCGTATCGTTTGTGGAGA 935924
QY 805 TCTGGAAGA 813
Db 935923 TATTGAATA 935915

RESULT 8
US-10-981-687-1/c
; Sequence 1, Application US/10981687
; Publication No. US20050131222A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; FILE REFERENCE: PB196P2C1D12
; CURRENT APPLICATION NUMBER: US/10/981,687
; PRIOR FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40809)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4416)..(4416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
```


NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature

LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
Query Match 34.5%; Score 319.4; DB 9; Length 1830121;
Best Local Similarity 65.8%; Fred. No. 4.4e-94;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
QY 85 AATTTAGGTAGCAAAATGATGAGACCCCGACATTTCTTATCGTTGAAGACGAGTTGGTAAAC 144
DB 936640 AGTATAGGATGGAAATGACTACTCTCCAAAATTTCTCGTTGTTGAAGATGAATTTGTAC 936581
QY 145 ACGCAACACGTTGAAAAGTATTTTCGAAGCGGAAGGCTATGATGATTTTCGAAGCGACAGA 204
DB 936580 TCGAAATACGCTTAAAGGGATTTTGAAGCGGAAGGATATGATGTTGTTGAAGCGAAAA 936521
QY 205 TGGCGCGGAATGATCAGATCCTCTCTGATATGACATCAACCTGGTGTGATCATGATAT 264
DB 936520 TGGTGTGAAATGATCATATATTGGCAAAATCAATAATTAATTTGGTTGTGATGATAT 936461
QY 265 CAATCTCGCGGTAAAGACGGTCTTCTGTAGCGGTGAACTGCGCGAGCAGGCGCAATGT 324
DB 936460 TAAATTACGAGCAAAAACGGCTTATTAATTGGCAAGAAATCACTCCGTGAAGATTAAGCTT 936401
QY 325 TGGTGTGATGTTCTGACTGCGCGGTGCAACGAAGTCAATAAATTTCTGGCCTCGAAAT 384
DB 936400 ACCTCTTATTTTAACTGTCGAGATAATGAAGTGGATAAAATTTTAGGTTTGGAAAT 936341
QY 385 CGGTGAGATGACTACATCACCAACGGTTCACACCGCGGTGAACTGACGATTCGTGACG 444
DB 936340 TGGTGGCGATGACTATTTGACTAAGCCTTTTAAACCTTAGAAGACTGACTATCCGCGCAG 936281
QY 445 CAACCTACTGTCGGTACCATGAATCTGGGTACTGTCTAGCGAAGAAACGTCGTAGCGTTGA 504
DB 936280 TAATCTATTGATCGTGAATG---CCGCATCAAGAAAAAGAAATACATTTGTCGAGA 936224
QY 505 AAGTACAAGTTCAATGTTGGGAATCTGGAATCAACACGCGCTGCTGTTGATGCGGCCCTGA 564
DB 936223 ATTCTATCGTTTAAATGTTGAAATTAGACTAAATAGCACTAAATAGCCACAGTTTAAATTA 936164
QY 565 TGGCGAGCAGTACAAAGTCCGCGCGAGCGAGTTCCGCGCCATGCTTCACTTCTGTGAAAA 624
DB 936163 AGGACAAAGATTCAAACTTCTCGCAGTGAATTTGTCGCAATGTTTACATTTCTGTGAAA 936104
QY 625 CCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAAGAAAAATGACCGCCGCTGAGCTGAA 684
DB 936103 TCCAGGAAAAATTCNAACGCGGAGAAATTCCTGAAAAAATGACCGGACGAGATTAA 936044
QY 685 ACCGACGACCGTACTGTAGACGTGACGATCCGCGCTATTCGTAACATTTTGAATCTAC 744
DB 936043 ACCTCAGGATCGTACCGTAGATGTCAAAATTCGAGTATCAGAAAAACATTTTGAAGATCA 935984
QY 745 GCGGATACGCGGAAATCATCGCCACCATTCACGGTGAAGTTATCGCTTCTGCGGTGA 804
DB 935983 TCCCAATACTCCAAATATCATTTAGCTATATCATGAGAGAGGCTATCGTTTTTGTGGAGA 935924
QY 805 TCTGGAAGA 813
DB 935923 TATTGAATA 935915
RESULT 9
US-09-815-242-7046
Sequence 7046, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.

```
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7046
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(711)
US-09-815-242-7046

Query Match      34.1%; Score 316.2; DB 3; Length 711;
Best Local Similarity 66.2%; Pred. No. 6.3e-95;
Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 101 ATGCAGACCCGCACATTTCTTATCGTTGAAGCAGAGTTGGTAACACGCAACACGTTGAAA 160
Db 1 ATGACTACTCCAAAATTTCTGGTTGAGATGAATTTGTCACTCGGAATACGCTTAAA 60

QY 161 AGTATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGCACAGATGGCGCGGAATGCA 220
Db 61 GGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCAGAAAATGGTGTGAAATGCA 120

QY 221 CAGATCCTCTGATATGACATCAACCTGGTGATCATGGATATCAATCTGCGGGTAAG 280
Db 121 CATATATTGGCAATCATTAATTAATTTGGTTGTGATGGAATTAATTTACCAGGCAAA 180

QY 281 AACGGTCTCTGTAGCGGTGAACCTGCGCAGCAGGCGAATGTTGCGTTGATGTTCTCG 340
Db 181 AACGGCTTATTTTGGCAGAGAACTCCGTGAAGATTAAGCTTACCTCTTATTTTAA 240

QY 341 ACTGCGCGTGACAAAGAGTCGATAAAATTTCTCGCCCTCGAAATCGGTGCAATCACTAC 400
Db 241 ACTGGTCGAGATAATGAAGTCGATAAAAATTTTAGGTTTGGAAATTTGGTGGCGGACTAT 300

QY 401 ATCACCACCCCTTCAACCCGCGTGAACCTGACGATTCGTGACGCAACCTACTGTCGGT 460
Db 301 TTGACTAAGCCCTTTAAACCCCTAGAGAACTGACTATCCGCGCAGCTAATCTATTGCA 360

QY 461 ACCATGAACTCGGTACTGTGACGCAAGAAACGTCGTAGCGTTGAAAGCTCAACAGTTCAAT 520
Db 361 GCATG- --CCGCATCAAGAAAAGAAAATACATTTGGTCGAGAATTTCTATGTTTAA 417

QY 521 GGTGGGAATCGACATCAACAGCCGTTCTGTTGATCGGCCCTCGATGGCGAGCAGTACAG 580
Db 418 GGTGGGAATTAGACCTAAATAGCCACAGTTTTAAATACACAGAAAGGACAGAAATTCAAA 477

QY 581 CTGCGCGCAGGAGTTCCGCGCCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAG 640
Db 478 CTTCTCGCAGTGAATTTCTGTCGAATGTTTACATTTCTGTGAAAATTCACGAGAAAATTC 537

QY 641 TCCCGTGCTGAACTGCTGAAGAAAATGACCGCGCTGAGCTGAAACCCGACCGACCGTACT 700
Db 61 GGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCAGAAAATGGTGTGGAATGCA 120

Db 538 ACGCGCAAGAAATTGCTGAAAAAATGACGCGCAGAGATTAAAACTCAGGATCGTACC 597
QY 701 GTAGACGTGAGATCCGCGGTATTCGTAACATTTTGAATCTACGCGGATACGCGGAA 760
Db 598 GTAGATGTCAATTCGACGATATCAGAAAACATTTTGAAGATCATCCCAATACTCCAAT 657
QY 761 ATCATCGCCACCATTCACGGTGAAGGTTATCGCTTCTGCGGTGATCTGGAAGA 813
Db 658 ATCATTATGACTATACATGGAAGGCTATCGTTTTTGTGGAGATATTGAATA 710

RESULT 10
US-10-282-122A-22157
; Sequence 22157, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22157
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-282-122A-22157

Query Match      34.1%; Score 316.2; DB 7; Length 711;
Best Local Similarity 66.2%; Pred. No. 6.3e-95;
Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 101 ATGCAGACCCGCACATTTCTTATCGTTGAAGCAGAGTTGGTAACACGCAACGTTGAAA 160
Db 1 ATGACTACTCCAAAATTTCTGGTTTGAAGATGAATTTGTCACTCGGAATACGCTTAAA 60

QY 161 AGTATTTTCGAAGCGGAAGGCTATGATGTTTTTCGAGGCGCACAGATGGCGCGGAATGCA 220
Db 61 GGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCAGAAAATGGTGTGGAATGCA 120
```



```
Db 357 GCTATCGCTTTTGGCGGAGATATTGAA 382

RESULT 13
US-10-795-159-230/c
; Sequence 230, Application US/10795159
; Publication No. US20050221439A1
; GENERAL INFORMATION:
; APPLICANT: BKALETZ et al.
; TITLE OF INVENTION: GENES OF AN OTITIS MEDIA ISOLATE OF NONTYPEABLE H. INFLUENZAE
; FILE REFERENCE: 28335/38815A
; CURRENT APPLICATION NUMBER: US/10/795.159
; CURRENT FILING DATE: 2004-03-05
; PRIOR APPLICATION NUMBER: US 60/453,134
; PRIOR FILING DATE: 2003-03-06
; NUMBER OF SEQ ID NOS: 771
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 230
; LENGTH: 2227
; TYPE: DNA
; ORGANISM: H. influenzae
US-10-795-159-230

Query Match 16.6%; Score 153.6; DB 9; Length 2227;
Best Local Similarity 66.8%; Pred. No. 6.3e-40;
Matches 219; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 85 AATTAGGTAGCAACATGCAGACCCCGCACATCTTATCGTTGAAGACGAGTTGGTAAC 144
Db 353 AGTATAAGGTGGAATAATGACTACTCCAAAATTTCTCGTTGTTGAAGATGAAATTTGTCAC 294
QY 145 AGCCAACAGCTTGAAGAATATTTTGAAGCGGAAGCTATGATGTTTTCGAAGCGACAGA 204
Db 293 TCGAATAACGCTTAAAGGATTTTGAAGCGGAAGGATATGATGTTTGAAGCGAAGAAA 234
QY 205 TGGCGGGAATGCATCAGATCCTCTCTGAATATGACATCAACCTGGTGTATCATGGATAT 264
Db 233 TGGTGTGAAATGCATATATATTTGGCAATCATATATTAATTTGGTTGTGATGATAT 174
QY 265 CAATCTGCGGTGAAGACGGTCTTCTGTAGCGGTGAACCTGCGGAGCAGGCGGAATGT 324
Db 173 TAAATTAACGAGCAAAAACGGCTTATTATTGGCAAGAGAACTCCCGTGAAGAAATTAAGCTT 114
QY 325 TCGGTGTATGTTCTGACTGCGGCGTGACAAACGAACGATCGATAAATCTCGGCTCGAAT 384
Db 113 ACCTCTTATTTTAACTGGTTCGAGATAATGAAGTGGATAAAATTTTAGGTCTGGAAT 54
QY 385 CGGTGCAGATGACTACATCACCACCAACCG 412
Db 53 TGGTCCGACGATTTATTAAACCAACCG 26

RESULT 14
US-10-770-824-1
; Sequence 1, Application US/10770824
; Publication No. US20040198954A1
; GENERAL INFORMATION:
; APPLICANT: Campos, Manuel
; APPLICANT: Baarsch, Mary Jo
; APPLICANT: Rosey, Everett
; APPLICANT: Ankenbauer, Robert
; APPLICANT: Warren-Stewart, Lynn
; APPLICANT: Suiter, Brian
; APPLICANT: Keach, Robin
; TITLE OF INVENTION: NOVEL PROTEINS FROM ACTINOBACILLUS PLEUROPNEUMONIAE
; FILE REFERENCE: PC9854A
; CURRENT APPLICATION NUMBER: US/10/770,824
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: US/09/418,980
; PRIOR FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
```

```
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Actinobacillus pleuropneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (272)..(787)
US-10-770-824-1

Query Match 11.6%; Score 107.4; DB 8; Length 1018;
Best Local Similarity 80.3%; Pred. No. 1.4e-24;
Matches 126; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 662 AAAATGACCGCGCGTGAAGCTGAAACCGCACACCGTACTGTAGAGCTGACGATCCGCGT 721
Db 68 AAAATGACCGCGTGTGAGCTAAACCGCATGACCGTACTGTGGATGTGACGATTCGCGT 127
QY 722 ATTGCTAAACATTTTGAATCTACCGCGGATACGCGGAAATCATCGCCACCATTCACCGT 781
Db 128 ATTGCTAAACACTTTGAAGATCACCTTAATACCGGAAATCATTTGTAACCATTCATGCT 187
QY 782 GAAGTTATCGCTTCTCGGCTGATCTGGAAGATTAAT 818
Db 188 GAAGGTTACCGTTTTTTCGCGCGAGTTAGTAGTAAT 224

RESULT 15
US-10-240-689-39/c
; Sequence 39, Application US/10240689
; Publication No. US2003017543A1
; GENERAL INFORMATION:
; APPLICANT: NG, Wee Chit
; TITLE OF INVENTION: Molecular Markers
; FILE REFERENCE: 6565-65001/RJP
; CURRENT APPLICATION NUMBER: US/10/240,689
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: SG 200002150-1
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 11275
; TYPE: DNA
; ORGANISM: E. coli
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: AE 000201
US-10-240-689-39

Query Match 11.5%; Score 106.8; DB 6; Length 11275;
Best Local Similarity 51.3%; Pred. No. 8.8e-24;
Matches 327; Conservative 0; Mismatches 302; Indels 9; Gaps 3;

QY 102 TGCAGACCCCGCACATTTCTATCGTTGAAGACGAGTTGGTAACACGCAACACGTTGAAA 161
Db 6729 TGATGCCACATCACATTTGTTTGAAGATGAGCCGGTTACCCAGGCGCGATTAACAAT 6670
QY 162 GTATTTTCGAAGCGGAAGCTATGATGTTTTCGAAGCGACAGATGCGCGGAAATGCATC 221
Db 6669 CCTACTTCACTCAGGAGGGGTATACCGTTTCGTTACGCGGAGCGGTCTGGGCTACGG 6610
QY 222 AGATCCTCTCTGAATATGACATCAACCTGGTGTATCATGATATCAATCTGCGGGGTAA 281
Db 6609 AAATTTATGCAATCATGTCGGTAGATTTAATTTCTGCTGGATATCACTTACCGATGAA 6550
QY 282 ACGGTCTTCTGTAGCGGTGAACCTGCGGAGCAGCGGAATGTTGGTTGATGTTCTTGA 341
Db 6549 ATGGCTGATGTTAAACCGCGCCCTGCGAGAACGCTCAACGCTGGGGATTTATTCGGTTA 6490
QY 342 CTGGCGGTGACAAACGAAGTCGATAAAATCTCGGCTCGAATCGTAATCGTCAGATGACT 401
Db 6489 CCGGACGCGAGCATCGGATGACCGTATTTGTTGGGCTGGAATGCGGCGACGACGATTAC 6430
QY 402 TCACCAACCGTTCAACCCCGGTGAACCTGATGACGATTCGTGTCAGCAACCTTACTGTCC 461
```

```
Db 6429 T C A C C A A A C C G C T G G A A C T G C C G A A C T G G T A G T A C G G G T G A A A A T C T G C T C G C G -- 6372
Qy 462 C C A T G A A T C T G G T A C T G T C A G C G A A G A A C G T C G T A G C G T T G A A A G C T A C A G T T C A A T G 521
Db 6371 - A A T G A C C T C G C G G A C A A G C T C A C C G C A C A C T C A G G A C A A C T G C T A T G C C T T T G C C G 6313
Qy 522 G T T G G G A A C T G A C A T C A A C A G C C G T T G T T G A T C G C C C T G A T G G C G A G C A G T A C A A G C 581
Db 6312 G T T A T T G C C T G A A T G T G T C G C C C A T A C G C T G -- G A G C G G A T G C G A G C C G A T T A A C 6256
Qy 582 T G C C G C G A G C G A G T T C C G C C A T G C T T C A C T T C T G T G A A A C C A G C G A A A T T C A G T 641
Db 6255 T G A C C C G C G C A G A T A T G A A A T G T T G T G G C A T T T G T G A C G A A T C C G G C G A A A T T C T C A 6196
Qy 642 C C C G T G C T G A A C T G C T G A A G A A A A T G A C C G C C G T G A G C T G -- A A A C G C A C G A C C G T A 698
Db 6195 G C C G T G A A C G T C T G T A C G T A T G C T T T G C C G C T C G G T G G A A A C C C T G A C C T G C G C A 6136
Qy 699 C T G T A G C G T G A C G A T C C G C C G T A T T C G T A A C A T T T C 736
Db 6135 C G T C G A T G T G T T A A T T C G T C G T T T A C G T C A T A A A C T C 6098
```

Search completed: February 13, 2006, 15:57:23
Job time : 847 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:34:44 ; Search time 203 Seconds
(without alignments)
515.134 Million cell updates/sec

Title: US-10-613-990A-32

Perfect score: 1230

Sequence: 1 MPTPHILIVEDLVTRNLK.....PEIIATIHGEGYRFGDLED 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A Geneseq_21.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*
- 9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1230	100.0	238	8	ADJ38403 Echerichia
2	1197	97.3	241	7	ABO62466 Klebsiell
3	1145	93.1	238	8	ADJ38391 Pantoea a
4	1138	92.5	239	6	ABM67466 Phototrab
5	1133	92.1	241	7	ADF05920 Bacterial
6	964.5	78.4	236	4	AU35550 Haemophil
7	964.5	78.4	236	6	ABU30417 Protein e
8	778	63.3	176	4	ABG28752 Novel hum
9	458	37.2	278	7	ABO84242 Pseudomon
10	400	32.5	236	5	ABP26003 Streptoco
11	400	32.5	236	6	ABU46510 Protein e
12	400	32.5	721	7	ABO79547 Pseudomon
13	394	32.0	236	5	ABP26002 Streptoco
14	394	32.0	236	8	ADV88454 Streptoco
15	394	32.0	236	8	ADV81866 Streptoco
16	394	32.0	236	8	ADV79707 Streptoco
17	389.5	31.7	237	5	ABB48823 Listeria
18	389.5	31.7	237	6	ABU32830 Protein e
19	389.5	31.7	240	6	ABM70299 Phototrab
20	386.5	31.4	235	6	ABU18458 Protein e
21	386.5	31.4	235	6	ABU44248 Protein e
22	384.5	31.3	249	7	ADF06436 Bacterial
23	376	30.6	233	4	AAU33417 Enterococ
24	376	30.6	234	4	AAU35155 Enterococ

ALIGNMENTS

RESULT 1

ADJ38403

ID ADJ38403 standard; protein; 238 AA.

XX AC ADJ38403;

XX AC ADJ38403;

XX DT 06-MAY-2004 (first entry)

XX DE Echerichia coli Arca (aerobic respiration control) protein SeqID32.

XX KW gamma-proteobacterium; Arca; aerobic respiration control;

XX KW Arca protein function; chromosomal arca gene; fermentation; L-amino acid;

XX KW L-lysine; L-glutamic acid.

XX OS Escherichia coli.

XX FN EP1382686-A1.

XX PD 21-JAN-2004.

XX PF 11-JUL-2003; 2003EP-00015911.

XX PR 12-JUL-2002; 2002JP-00203764.

XX (AJIN) AJINOMOTO CO INC.

XX Ishikawa Y, Imaizumi A, Matsui K, Kojima H;

XX WPI; 2004-192975/19.

XX N-PSDB; ADJ38402.

XX New bacterium useful for producing a target substance (especially L-

XX lysine, L-glutamic acid or another L-amino acid) is modified so that the

XX Arca protein does not normally function.

XX Claim 2; SEQ ID NO 32; 38pp; English.

XX This invention relates to a novel gamma-proteobacterium able to produce a

XX target substance such as an amino acid which is modified so that an Arca

XX (aerobic respiration control) protein does not normally function. The

XX disruption to Arca protein function is preferably due to disruption of a

XX chromosomal Arca gene. The bacterium and method are useful in producing a

XX target substance by fermentation. The target substance is preferably an L

XX -amino acid, especially L-lysine or L-glutamic acid. Production of the

XX target substance using the modified bacterium is more efficient. The

XX present sequence is that of the E coli Arca protein which was used in the

XX method of the invention.

XX	SQ	Sequence 238 AA;	Query Match	100.0%;	Score 1230;	DB 8;	Length 238;
			Best Local Similarity	100.0%;	Pred. No. 2.6e-127;		
			Matches 238;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MQPHILIVDELVTRNTLKSIFPAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP	60				
Db	1	MQPHILIVDELVTRNTLKSIFPAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP	60				
Qy	61	NGLLARELREQANVALMELTGRDNEVDKILGLEIGADDYITKGFNPRELTIARNLLSR	120				
Db	61	NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKGFNPRELTIARNLLSR	120				
Qy	121	TMNLTGVSEERRSVESYKFGNWELDINSRSLIGPDGEQYKLPSEFRAMLHFCENPGKIQ	180				
Db	121	TMNLTGVSEERRSVESYKFGNWELDINSRSLIGPDGEQYKLPSEFRAMLHFCENPGKIQ	180				
Qy	181	SRAELLKQMTGRELKPDRHTVDVTVIRIRKHFEPTPTPEIATIHGSGYVFCGDLED	238				
Db	181	SRAELLKQMTGRELKPDRHTVDVTVIRIRKHFEPTPTPEIATIHGSGYVFCGDLED	238				

RESULT 2
ABO62466
ID ABO62466 standard; protein; 241 AA.
XX
XX ABO62466;
XX
XX
XX 29-JUL-2004 (first entry)
XX
XX
XX Klebsiella pneumoniae polypeptide secid 8903.
DE

Query Match 97.3%; Score 1197; DB 7; Length 241;
 Best Local Similarity 97.1%; Pred. NO. 1.2e-123;
 Matches 231; Conservative 5; Mismatches 2; Indels 0; Gaps 0

Qy	1	MQPHILIVDELVTRNTLXSIPEAEGYDVPEATDGAEMHQILSEYDINLVIMDINLP	60
Db	4	MQPHILIVDELVTRNTLXSIPEAEGYDVPEATDGAEMHQILSENINLVIMDINLP	63
Qy	61	NGLLARELREQANVALMFUTGRDNEVDKILGLEIGADDYTKPNPRELTIARANLUS	120
Db	64	NGLLARELREQADVALMFUTGRDNEVDKILGLEIGADDYTKPNPRELTIARANLUS	123
Qy	121	TNMLGTVSEBRRSVESYKFGNWELDINRSLSIGPDGEQYKLPSEPRAMLHFCENPGKIQ	180
Db	124	TNMLGTVSEBRRSVESYKFGNWELDINRSLSVSPNGEQYKLPSEPRAMLHFCENPGKIQ	183
Qy	181	SRAELLKXMTGRELKPHDRTVDTVIRIRKHFESTPTDPTPEIIATHGBGYFCGDLED	238
Db	184	SRAELLKXMTGRELKPHDRTVDTVIRIRKHFESTPTDPTPEIIATHGBGYFCGDLOE	241

RESULT 3	
ADJ38391	
ID	ADJ38391 standard; protein; 238 AA.
XX	
XX	
AC	ADJ38391;
XX	
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Pantoea ananatis Arca (aerobic respiration control) protein seqID20.
XX	
KW	gamma-proteobacterium; Arca; aerobic respiration control;
KW	Arca protein function; chromosomal arca gene; fermentation; L-amino acid;
KW	L-lysine; L-glutamic acid.
XX	
XX	
OS	Pantoea ananatis.
XX	
PN	EPI382686-A1.
XX	
PD	21-JAN-2004.
XX	
PF	11-JUL-2003; 2003EP-00015911.
XX	
PR	12-JUL-2002; 2002JP-00203764.
XX	
XX	
PA	(AJIN) AJINOMOTO CO INC.
XX	
PI	Ishikawa Y, Imaizumi A, Matsui K, Kojima H;
XX	
DR	WPI; 2004-192975/19.
DR	N-PSDB: ADJ38390.

Example 4; SEQ ID NO 20; 38pp; English.

This invention relates to a novel gamma-proteobacterium able to produce a target substance such as an amino acid which is modified so that an ArCA (aerobic respiration control) protein does not normally function. The disruption to ArCA protein function is preferably due to disruption of a chromosomal arca gene. The bacterium and method are useful in producing a target substance by fermentation. The target substance is preferably an L-amino acid, especially L-lysine or L-glutamic acid. Production of the target substance using the modified bacterium is more efficient. The present sequence is that of the P ananatis Arca protein which was used in the exemplification of the invention.

Sequence 238 AA;

Query Match 93.1%; Score 1145; DB 8; Length 238;
Best Local Similarity 92.0%; Pred. No. 6.8e-119;
Matches 219; Conservative 11; Mismatches 8; Indels 0; Gaps 0

Db 1 MQTPHILIVEDLVTRNTLKSIFEAAGYVVVEATDGAEMHQVLTDNDVNLVIMDINLPKG 60
QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIARNLISR 120
Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIARNLISR 120
QY 121 TNLGTVSEERSVSYPKNGWELDINSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180
Db 121 TNLPLPNEERRQVESYKFNWELDINSRLSNPNNGEQYKLPSEFRAMLHFCENPGKIQ 180
QY 181 SRAELKKMTGRELKHDPHRTVDVTIRIRKHFPSTPDPTPEIIATIHGEGYRFGCDLE 238
Db 181 TRADLLKMTGRELKHDPHRTVDVTIRIRKHFPSTPDPTPEIIATIHGEGYRFGCDLQ 238

RESULT 4

ARM67466
ID ARM67466 standard; protein; 239 AA.

AC ARM67466;

DT 20-NOV-2003 (first entry)

DE Photorhabdus luminescens protein sequence #563.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.

XX Photorhabdus luminescens.

OS WO200294867-A2.

PN 28-NOV-2002.

PD 07-FEB-2002; 2002WO-IB003040.

PF 07-FEB-2001; 2001FR-00001659.

PR (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Tacurit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;

PI WPI; 2003-148459/14.

DR Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.

PS Claim 2; SEQ ID NO 563; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 239 AA;

Query Match 92.5%; Score 1138; DB 6; Length 239;

Best Local Similarity 93.2%; Pred. No. 4.1e-117;

Matches 221; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAAGYVDVEATDGAEMHQILSEYDINLVIMDINLPKG 60

Db 1 MQTPHILIVEDLVTRNTLKSIFEAAGYIVVEATDGEHMLSNNDINLVIMDINLPKG 60

QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIARNLISR 120

Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIARNLISR 120

QY 121 TNLGTVSEERSVSYPKNGWELDINSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180

Db 121 TNLNSVSEERRQVESYKFNWELDINSRLSLSPAGEPYKLPSEFRAMLHFCENPGKIQ 180

QY 181 SRAELKKMTGRELKHDPHRTVDVTIRIRKHFPSTPDPTPEIIATIHGEGYRFGCDLE 237

Db 181 TRADLLKMTGRELKHDPHRTVDVTIRIRKHFPSTPDPTPEIIATIHGEGYRFGCDLE 237

RESULT 5

ADP05920

ID ADP05920 standard; protein; 241 AA.

AC ADP05920;

DT 12-FEB-2004 (first entry)

DE Bacterial polypeptide #2033.

XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant.

OS Proteus mirabilis.

PN US6605709-B1.

PD 12-AUG-2003.

PF 05-APR-2000; 2000US-00543681.

PR 09-APR-1999; 99US-0128706P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Breton GL;

XX WPI; 2003-895291/82.

DR N-PSDB; ADP01748.

XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.

XX Disclosure; SEQ ID NO 6205; 870pp; English.

XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This

CC sequence represents a *Proteus mirabilis* polypeptide of the invention.
 XX
 SQ Sequence 241 AA;
 Query Match 92.1%; Score 1133; DB 7; Length 241;
 Best Local Similarity 92.0%; Pred. No. 1.5e-116;
 Matches 219; Conservative 8; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVPEATDGAEMHOILSEYDINLVMDINLPCK 60
 DB 4 MQTPHILIVEDLVTRNTLKSIFEAEGYDVPEATDGAEMHOILSEYDINLVMDINLPCK 63
 QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 DB 64 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 123
 QY 121 TNNLGTSEERRSVSYKFNWELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKI 180
 DB 124 TNNLANGTEERLVSYSKFNWELDINSRLISPTGEQYKLPFRSEFRAMLHFCENPGKI 183
 QY 181 SBAELLKQMTGRELPHDRTVDVTTIRIRKHPFESTPDTEIATIHGGYRFGDLE 238
 DB 184 TRAEELLKQMTGRELPHDRTVDVTTIRIRKHPFESTPDTEIATIHGGYRFGDLE 241
 RESULT 6
 AAU3550
 ID AAU3550 standard; protein; 236 AA.
 AC AAU3550;
 XX
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Haemophilus influenzae cellular proliferation protein #191.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 XX Haemophilus influenzae.
 XX
 XX WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US009180.
 XX
 PR 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS53409.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 11143; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also

CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: the sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 236 AA;
 Query Match 78.4%; Score 964.5; DB 4; Length 236;
 Best Local Similarity 76.2%; Pred. No. 6.9e-98;
 Matches 183; Conservative 26; Mismatches 24; Indels 7; Gaps 2;
 QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVPEATDGAEMHOILSEYDINLVMDINLPCK 60
 DB 1 MTTPKILVEDEIVTRNTLKGIFEAEGYDVPEAENGVMHILAHNINLVMDINLPCK 60
 QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 QY 121 TNNLGTSEERRSV---ESYKFNWELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPG 177
 DB 121 AM---PHQEKENTFGREYRFGNWKLDLNSHSLITPEGQFKLPFRSEFRAMLHFCENPG 176
 QY 178 KIOSRAELLKQMTGRELPHDRTVDVTTIRIRKHPFESTPDTEIATIHGGYRFGDLE 237
 DB 177 KLOTREELLKQMTGRELPHDRTVDVTTIRIRKHPFESTPDTEIATIHGGYRFGDIE 236
 RESULT 7
 ABU30417
 ID ABU30417 standard; protein; 236 AA.
 XX
 AC ABU30417;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #15944.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Haemophilus influenzae.
 XX
 FN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA34287.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
KW Pseudomonas aeruginosa.
OS US6551795-B1.
XX 22-APR-2003.
PD 18-FEB-1999; 98US-00252991.
XX 18-FEB-1998; 98US-0074788P.
PF 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PA Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
PI N-PSDB; ABD17813.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX Disclosure; SEQ ID NO 32988; 455pp; English.
PS The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and the
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67826-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX Sequence 278 AA;
SQ

Query Match 37.2%; Score 458; DB 7; Length 278;
Best Local Similarity 39.4%; Pred. No. 1.3e-41;
Matches 95; Conservative 47; Mismatches 91; Indels 8; Gaps 3;
QY 4 PHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSE-----YDINLVINDINLP 58
DB 39 PRVLVVDVDPVIRELLQAYLGEYDVLCAAGNAEQAEACLAECALHGPVELVLDIRLP 98
QY 59 GRNGLLARELREQANVLMFTGRDNEVDKTLGLEIGADDYITKPFNPRTIARNLL 118
DB 99 GGGTLTRELAVRSEVGILLTGRNDEIDRLVGLGCGADDYIVKPLNPRELVSAKLI 158
QY 119 SRTMNL-GTVSEERSVSYSKFNKGWELDINSRLIGPDGEQYKLPSEFRAMLHFCENPG 177
DB 159 RRVVHAQASAGPARQALR--QFGDMLLDADRRRLIDHAGNETLLTHGEFQLGAFLRNSG 216
QY 178 KIQSAELLKQMTGRELKPHDRTVDVTTIRIRKHFESTPDPTPEIATTHGEGYRCGDLE 237
DB 217 HTLSRDQLMDQIRNRELWFLPSDRSIDVLVGLRRRLKRDPAEPQLIITTHGAGYLFATAAS 276
QY 238 D 238
DB 277 D 277

RESULT 10
ABP26003

ID ABP26003 standard; protein; 236 AA.
XX
AC ABP26003;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 1182.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus pyogenes.
XX
XX WO200234771-A2.
XX
PD 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
PF
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN66634.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
PS Claim 1; Page 3273; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 236 AA;

Query Match 32.5%; Score 400; DB 5; Length 236;
Best Local Similarity 39.5%; Pred. No. 2.7e-35;
Matches 92; Conservative 45; Mismatches 82; Indels 14; Gaps 4;
QY 6 ILIIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVINDINLPKNGLL 65
DB 4 ILIIVDEKPISDIIFKFNLTKEGYDITVAFDGREAVTIFEEKPDLIILDMLPELDGLEV 63
QY 66 ARELREQANVLMFTGRDNEVDKTLGLEIGADDYITKPFNPRTIARNLLSRTMNLG 125
DB 64 AREIKTSHVPIIMLSAKDSFVKIGLEIGADDYITKPFNRELLARVKAHLRTETIE 123
QY 126 T-VSEERSVSYSKFNKGWELDINSRLIGPD-----GEQYKLPSEFRAMLHFCENPG 177

Db 124 TAVAEENASSGTQ-----ELTIGNQLIL-PDAFVAKKHQGEVELTHREFELHLHANMG 177

Qy 178 KIQSRAELLKKMTGRELKHDPHRTVDVTTIRRKHFESTPDPPEIATIHGEGY 230

Db 178 QVMTREHLLLEIWMGYDYFGDVRTVDVTVRLREKIEDTPSRPEXILTRRGVGY 230

RESULT 11

ABU46510

ID ABU46510 standard; protein; 236 AA.

XX AC ABU46510;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #32037.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Streptococcus pyogenes.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072831.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACA50380.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 74434; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

XX proliferation; (7) identifying a compound that influences the activity of

XX the gene product or that has an activity against a biological pathway

XX required for proliferation, or that inhibits cellular proliferation; (8)

XX identifying a gene required for cellular proliferation or the biological

XX pathway in which a proliferation-required gene or its gene product lies

XX or a gene on which the test compound that inhibits proliferation of an

XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX compound's activity; (11) a culture comprising strains in which the gene

XX product is overexpressed or underexpressed; (12) determining the extent

XX to which each of the strains is present in a culture or collection of

XX strains; or (13) identifying the target of a compound that inhibits the

XX proliferation of an organism. The antisense nucleic acids are useful for

XX identifying proteins or screening for homologous nucleic acids required

XX for cellular proliferation to isolate candidate molecules for rational

XX drug discovery programs, or for screening homologous nucleic acids

XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 236 AA;

Query Match 32.5%; Score 400; DB 6; Length 236;

Best Local Similarity 39.5%; Pred. No. 2.7e-35;

Matches 92; Conservative 45; Mismatches 82; Indels 14; Gaps 4;

Qy 6 ILIVEDELVTNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVINDINLPGKNGLL 65

Db 4 ILIVDEKPISDIIKFNLTKEGYDITVAFDGREAVTIFESEKPDIIILDLMLPELDGLEV 63

Qy 66 ARELREQANVALMPLTGRDNEVDKILGLEIGADDYITKPFNPRLTIBARNLISMTNIG 125

Db 64 AREIRKTSHPITMILSAKSEDFKVGLEIGADDYITKPFNPRLTIBARNLISMTNIG 123

Qy 126 T-VSEERRSVESYKFNKGWELDINSRLIGPD-----GEQYKLPSEFRAMLHFCENPG 177

Db 124 TAVAEENASSGTQ-----ELTIGNQLIL-PDAFVAKKHQGEVELTHREFELHLHANMG 177

Qy 178 KIQSRAELLKKMTGRELKHDPHRTVDVTTIRRKHFESTPDPPEIATIHGEGY 230

Db 178 QVMTREHLLLEIWMGYDYFGDVRTVDVTVRLREKIEDTPSRPEXILTRRGVGY 230

RESULT 12

ABO79547

ID ABO79547 standard; protein; 721 AA.

XX AC ABO79547;

XX 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polypeptide #11722.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

XX 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

XX N-PSDB; ABD13118.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, if

XX useful as molecular targets for diagnostics, prophylaxis and treatment of

XX pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 28293; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the

XX polynucleotides encoding them. The sequences are useful in diagnosis and

XX therapy of pathological conditions, as molecular targets for diagnostics, and

XX prophylaxis and treatment of pathological conditions resulting from a

XX bacterial infection, for evaluating a compound, such as a polypeptide,

XX for the ability to bind a *P. aeruginosa* nucleic acid, as components of

XX effective antibacterial targets, as targets for antibacterial drugs,

XX including anti-*P. aeruginosa* drugs, as templates for recombinant

CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences AB067826-
CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html

XX Sequence 721 AA;

Query Match 32.5%; Score 400; DB 7; Length 721;
Best Local Similarity 38.2%; Pred. No. 1.5e-34;
Matches 89; Conservative 44; Mismatches 92; Indels 8; Gaps 3;

QY 5 HILIVEDLVTNLTLSIFEAEGYDVFEATDGAEMHQILSEYDINLVMIDNLPKNGLL 64
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
481 HILIVDDDRREITELVGNVILKKNGLRTTIVADGQVRAFLVLEANTVDLIVLDMPPGDDGLL 540
QY 65 LARELREQANVA--LMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSRTM 122
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
541 LCRELVGKHKATPVLMLTARNDETDRIGLEMGADDYITKPFNSARELLARINAVLRTR 600
QY 123 NLG---TVSEERRSVESYKFNKGWELDINSRLIGPDGEQYKLPKPRSEFRAMLHFCENPGKI 179
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
601 MLPNNLTVSESRLLI---GFGQWQLDTSARHLDDAGTVALSGAEYLLRVFLDHPQV 657
QY 180 QSRALLKQMTGRELKHDPDRTVDVTRIRKHFESTPDTPPEIATIHGEGYRF 232
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
658 LSRDQLNLQTGREADIFDRSIDLLVRLRLQRLGDDAREPEYIKTVRSEGYVF 710

RESULT 13

ABP26002
ID ABP26002 standard; protein; 236 AA.

XX AC ABP26002;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 1180.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

XX 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;

XX Tettelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN66633.

XX New Streptococcus protein for the treatment or prevention of infection or

XX disease caused by Streptococcus bacteria, such as meningitis, and for

XX detecting a compound that binds to the protein.

XX Claim 1; Page 3273; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX Sequence 236 AA;

Query Match 32.0%; Score 394; DB 5; Length 236;
Best Local Similarity 39.6%; Pred. No. 1.2e-34;
Matches 90; Conservative 44; Mismatches 91; Indels 2; Gaps 2;

QY 6 ILIVEDLVTNLTLSIFEAEGYDVFEATDGAEMHQILSEYDINLVMIDNLPKNGLL 65
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
4 ILIVDDKPSIDIIKFNLTKEGYETATAFDGRAALVQVAEPQDILILDLMLPELDGLEV 63
QY 66 ARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSRTM 125
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
64 AKEVKTSHIPIIMLSAKDSEFDKIVGLEIGADDYITKPFNSRELLARVKAHLRRTENIE 123
QY 126 T-VSEERRSVESYKFNKGWELDINSRLIGPD-GEQYKLPKPRSEFRAMLHFCENPGKIQSRA 183
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
124 TAVAESQANASSDITIGELQLLPDAFTAKRGREIBELTHREFLLHHLATHIGQVMTRE 183
QY 184 ELLKQMTGRELKHDPDRTVDVTRIRKHFESTPDTPPEIATIHGEGY 230
Dl |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
184 HLETVNGDYFGDVRTVTRIRKIEKIEDTPGRPEYILTRRGVGY 230

RESULT 14

ADV88454

ID ADV88454 standard; protein; 236 AA.

XX AC ADV88454;

XX 24-FEB-2005 (first entry)

XX Streptococcus agalactiae protein sequence, SEQ ID 848.

XX Antibacterial; Vaccine; bacterial infection.

XX Streptococcus agalactiae.

XX FR2824074-A1.

XX 31-OCT-2002.

XX 26-APR-2001; 2001FR-00005642.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier P, Frangeul L, Lalioui L;

XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

XX agalactiae for the development of vaccines, diagnostic tools, DNA chips

PT and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 848; 2687pp; French.

PS The present invention relates to novel Streptococcus agalactiae

XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;

CC agalactiae involved in the synthesis of amino acids, cell membranes,

CC intermediate (central) metabolism, energetic metabolism, fatty acid and

CC phospholipid metabolism, nucleotide metabolism including purines,

CC pyrimidines and/or nucleosides, regulatory functions, replication,

CC transcription, translation, protein transport, adaptation to atypical

CC conditions, sensitivity to medicines and/or analogues, functions related

CC to transporters, biosynthesis of cofactors, prosthetic groups and

CC transporters, cell membrane proteins and cellular machinery. (I) are

CC useful for the detection and/or amplification of nucleic acids.

CC Pharmaceutical composition comprising (I) or (II) are useful for

CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2

CC contains 6617 sequence whereas the present patent only contains 2344

XX sequences.

XX Sequence 236 AA;

SQ

Query Match 32.0%; Score 394; DB 8; Length 236;

Best Local Similarity 39.6%; Pred. No. 1.2e-34;

Matches 90; Conservative 44; Mismatches 91; Indels 2; Gaps 2;

QY 6 ILIIVDELVTNTLSIFEAGYDVFEATDCAEMHQILSEYDINLVIMDINLPKNGILL 65

DB 4 ILIIVDEKPISDIIKFNLTKEGYETATAFDGREALVQVAFQPDIIILDLMLPELDGLEV 63

QY 66 ARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARLLSRMTNLG 125

DB 64 AKEVRKTSHPILMILSAKSEFDKVGILEIGADDYITKPFNSRELLARVKAHLRRTENIE 123

QY 126 T-VSEERSVSYKFNWGHELDINSRLIGPD-GEQYKLPSEFRAMLHFCENPGKIQSRA 183

DB 124 TAVAESAQNASDITIGELQILPDATFAKKGEBIEITHREFELHHLATHIGQVMTRE 183

QY 184 ELLKMTGRELKPHDRTVDVTRIRKHFESTPTDPEIIATHIGEGY 230

DB 184 HLLTVMGYDYGVDVTRVTRRLEKIEDTPGRPEYILTRRGVGY 230

RESULT 15

ADV81866

ID ADV81866 standard; protein; 236 AA.

AC ADV81866;

XX

XX 24-FEB-2005 (first entry)

XX

DE Streptococcus agalactiae protein, SEQ ID 3007.

XX

XX Antibacterial; vaccine; bacterial infection.

XX

OS Streptococcus agalactiae.

XX

FN WO200292818-A2.

XX

PD 21-NOV-2002.

XX

PF 26-APR-2002; 2002WO-IB003059.

XX

PR 26-APR-2001; 2001FR-00005642.

XX

XX (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX

XX Glaser P, Rueniok C, Chevalier F, Frangeul L, Lalloui L;

PI Zouine M, Couve E, Buchrieser C, Foyart C, Trieu-Cuot P, Kunst F;

XX

DR WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips

XX and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 3007; 439pp; French.

PS The present invention relates to novel Streptococcus agalactiae

CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and

CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The

CC nucleotide sequences encode polypeptides of S. agalactiae involved in the

CC synthesis of amino acids, cell membranes, intermediate (central)

CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,

CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,

CC regulatory functions, replication, transcription, translation, protein

CC transport, adaptation to atypical conditions, sensitivity to medicines

CC and/or analogues, functions related to transporters, biosynthesis of

CC cofactors, prosthetic groups and transporters, cell membrane proteins and

CC cellular machinery. (I) are useful for the detection and/or amplification

CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are

CC useful for treatment of a bacterial S. agalactiae infection. The complete

CC genome of Streptococcus agalactiae is given in ADV81204. Note: The

CC present patent is an equivalent for the basic patent FR2824074A1, which

XX contains only 2344 sequences.

SQ Sequence 236 AA;

Query Match 32.0%; Score 394; DB 8; Length 236;

Best Local Similarity 39.6%; Pred. No. 1.2e-34;

Matches 90; Conservative 44; Mismatches 91; Indels 2; Gaps 2;

QY 6 ILIIVDELVTNTLSIFEAGYDVFEATDCAEMHQILSEYDINLVIMDINLPKNGILL 65

DB 4 ILIIVDEKPISDIIKFNLTKEGYETATAFDGREALVQVAFQPDIIILDLMLPELDGLEV 63

QY 66 ARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARLLSRMTNLG 125

DB 64 AKEVRKTSHPILMILSAKSEFDKVGILEIGADDYITKPFNSRELLARVKAHLRRTENIE 123

QY 126 T-VSEERSVSYKFNWGHELDINSRLIGPD-GEQYKLPSEFRAMLHFCENPGKIQSRA 183

DB 124 TAVAESAQNASDITIGELQILPDATFAKKGEBIEITHREFELHHLATHIGQVMTRE 183

QY 184 ELLKMTGRELKPHDRTVDVTRIRKHFESTPTDPEIIATHIGEGY 230

DB 184 HLLTVMGYDYGVDVTRVTRRLEKIEDTPGRPEYILTRRGVGY 230

Search completed: February 11, 2006, 02:38:20

Job time : 207 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:38:39 ; Search time 44 Seconds
(without alignments)
520.445 Million cell updates/sec

Title: US-10-613-990A-32
Perfect score: 1230
Sequence: 1 MQTPHILIVEDLVTRNLK.....PEIIATHGEGYFCGDLED 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1230	100.0	238	1 JYECR	dye resistance pro
2	1230	100.0	238	2 G91298	dye resistance pro
3	1230	100.0	238	2 A86140	dye resistance pro
4	1227	99.8	238	2 AF1076	global response re
5	1145	93.1	238	2 AG0056	aerobic respiratio
6	1098	89.3	238	2 H82084	aerobic respiratio
7	964.5	78.4	236	2 I64099	aerobic respiratio
8	535	43.5	234	2 G82164	DNA-binding respon
9	479	38.9	230	2 A64841	torCAD operon tran
10	476	38.7	230	2 B85635	response transcrip
11	476	38.7	230	2 F90772	response transcrip
12	458	37.2	244	2 H83022	probable two-compo
13	427	34.7	238	2 A33608	two component resp
14	420	34.1	238	1 S28674	hypothetical prote
15	400	32.5	246	2 F83133	probable two-compo
16	398	32.4	245	2 D95868	probable response
17	392	31.9	238	2 T31680	back protein - Bac
18	389.5	31.7	237	2 AH1110	two-component resp
19	384.5	31.3	237	2 AD1472	two-component resp
20	384	31.2	247	2 D72228	response regulator
21	379	30.8	236	2 C84153	two-component resp
22	378.5	30.8	232	2 A83794	two-component resp
23	378	30.7	254	2 AE0017	transcription regu
24	378	30.7	254	2 B87396	DNA-binding respon
25	375	30.5	235	2 B89760	response regulator
26	374	30.4	235	2 S85967	osmosensor respons
27	373.5	30.4	239	1 RGEOR	two-component resp
28	373.5	30.4	239	2 AD0998	two-component resp
29	373.5	30.4	239	2 G91159	response regulator

30	373.5	30.4	239	2 F86005	hypothetical prote
31	373.5	30.4	239	2 S01366	regulatory protein
32	373.5	30.4	253	2 AH3249	two component resp
33	373.5	30.4	253	2 S03760	virG protein - Agr
34	373	30.3	253	2 AG2368	two-component resp
35	372.5	30.3	239	2 S41887	ompR protein - Sal
36	371.5	30.2	245	2 AH0481	probable response
37	370	30.1	242	2 B83246	two-component resp
38	369.5	30.0	239	2 S70205	osmosensor respons
39	368	29.9	239	2 T06976	regulatory protein
40	367.5	29.9	241	2 JS0239	virG protein - Agr
41	365.5	29.7	267	2 A25519	virG protein - Agr
42	365	29.7	234	2 D95142	DNA-binding respon
43	365	29.7	251	2 B98010	response regulator
44	363.5	29.6	224	2 B97200	response regulator
45	358.5	29.1	240	2 C82042	transcription regu

ALIGNMENTS

RESULT 1

JYECR

dye resistance protein - Escherichia coli (strain K-12)

N/Alternate names: aerobic respiration control protein ArcA; protein dye negative regula

C/Species: Escherichia coli

C/Date: 28-May-1986 #sequence revision 28-May-1986 #text_change 09-Jul-2004

C/Accession: A03561; S56625; H65255

R/Drury, L.S.; Buxton, R.S.

J. Biol. Chem. 260, 4236-4242, 1985

A/Title: DNA sequence analysis of the dye gene of Escherichia coli reveals amino acid ho

A/Reference number: A92515; MUID:85157583; PMID:2984198

A/Accession: A03561

A/Molecule type: DNA

A/Residues: 1-238 <DR>

A/Cross-references: UNIPROT:P03026; UNIPARC:UPI0000125D8C; GB:M10044; NID:gl45817; PIDN:

R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995

A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.

A/Reference number: S56314; MUID:95334362; PMID:7610040

A/Accession: S56625

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-238 <BUR>

A/Cross-references: UNIPARC:UPI0000125D8C; EMBL:U14003; NID:gl263172; PIDN:AAA97297.1; F

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: H65255

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-238 <BLAT>

A/Cross-references: UNIPARC:UPI0000125D8C; GB:AE000510; GB:U00096; UMGP:b4401; NID:gl790

A/Experimental source: strain K-12, substrain MGL655

C/Genetics:

A/Gene: arcA; dye; fexA; msp; seg; sfrA

A/Map position: 100 min

C/Function:

A/Note: may also be involved in the osmoregulation of envelope proteins

C/Superfamily: ompR protein; response regulator homology

C/Keywords: DNA binding; osmoregulation; phosphoprotein; transcription regulation

F:6-114/Domain: response regulator homology <RRH>

F:54/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 100.0% Score 1230; DB 1; Length 238;

Best Local Similarity 100.0%; Pred. No. 1.1e-85;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQTPHILIVEDLVTRNLKSIFFAEGYDVFEATDGAEMHQITLSEYDINLVIMDINLPCK 60

|||||
Db 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
QY 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
Db 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
Db 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
RESULT 2
G91298
dye resistance protein arcA [similarity] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: G91298
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: G91298
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <HAV>
A:Cross-references: UNIPROT:P03026; UNIPARC:UPI0000125D8C; GB:BA0000007; PIDN:BAB38782.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC85359
C:Superfamily: ompR protein; response regulator homology
Query Match 100.0%; Score 1230; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.1e-89;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
Db 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
QY 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
Db 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
Db 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
RESULT 3
A86140
dye resistance protein arcA [similarity] - Escherichia coli (strain O157:H7, substrain B
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A86140
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: A86140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <STO>

A:Cross-references: UNIPROT:P03026; UNIPARC:UPI0000125D8C; GB:AE005174; NID:g12519433; F
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: arcA
C:Superfamily: ompR protein; response regulator homology
Query Match 100.0%; Score 1230; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.1e-89;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
Db 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
QY 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
Db 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
Db 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
RESULT 4
AF1076
global response regulator arcA [imported] - Salmonella enterica subsp. enterica serovar
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AF1076
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
. S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:21534947; PMID:11677608
A:Accession: AF1076
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <PAR>
A:Cross-references: UNIPARC:UPI000005A9F8; GB:AL513382; PIDN:CAD03429.1; PID:g16505695;
C:Genetics:
A:Gene: arcA
C:Superfamily: ompR protein; response regulator homology
Query Match 99.8%; Score 1227; DB 2; Length 238;
Best Local Similarity 99.6%; Pred. No. 1.8e-89;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
Db 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
Db 61 NGLLLARELREQANVALMFLTGRDNEVDKILGLEIGADDDYTKPFPNPRELTIARNLISR 120
QY 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
Db 121 TNNLGTVSEERSVSYKFNWGELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
Db 181 SRAELLKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGCGLED 238
RESULT 5
AG0056

aerobic respiration control protein [imported] - Yersinia pestis (strain C092)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AG0056
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AG0056
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-238 <KUR>
 A:Cross-references: UNIPROT:Q8ZIN7; UNIPARC:UPI00000DCC6B; GB:AL590842; PIDN:CAC89314.1;
 C:Genetics:
 A:Gene: arca
 C:Superfamily: ompR protein; response regulator homology

Query Match 93.1%; Score 1145; DB 2; Length 238;
 Best Local Similarity 92.9%; Pred. No. 5.3e-83;
 Matches 221; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATGAEMHQILSYDINLVIMDINLPKG 60
 Db 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATGAEMHQILSYDINLVIMDINLPKG 60

Qy 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Db 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120

Qy 121 TMNLGTVSEERSVESYKFNQWELDINSRSLIGPDGEQYKLPFRSEFRAMLHFCENPGKI 180
 Db 121 TMNLGTVSEERSVESYKFNQWELDINSRSLIGPDGEQYKLPFRSEFRAMLHFCENPGKI 180

Qy 181 SRABELLKQMTGRELKPQDRTVDVTRIRKHFESTPDTPPEIATIHGEGYRFGCDLE 238
 Db 181 SRABELLKQMTGRELKPQDRTVDVTRIRKHFESTPDTPPEIATIHGEGYRFGCDLE 238

RESULT 6
 H82084
 aerobic respiration control protein FexA VC2368 [imported] - Vibrio cholerae (strain N16
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C:Accession: H82084
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, E.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: AB2035; MUID:20406833; PMID:10952301
 A:Accession: H82084
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-238 <HEI>
 A:Cross-references: UNIPROT:Q9KJ9; UNIPARC:UPI00000C329F; GB:AE004307; GB:AE003852; NID
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC2368
 A:Map position: 1
 C:Superfamily: ompR protein; response regulator homology

Query Match 89.3%; Score 1098; DB 2; Length 238;
 Best Local Similarity 88.2%; Pred. No. 2.7e-79;
 Matches 210; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

Qy 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATGAEMHQILSYDINLVIMDINLPKG 60
 Db 1 MQTQILLIVEDQVTRNTLKSIFEAEGYDVFEATGAEMHQILSYDINLVIMDINLPKG 60

Qy 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120

Db 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Qy 121 TMNLGTVSEERSVESYKFNQWELDINSRSLIGPDGEQYKLPFRSEFRAMLHFCENPGKI 180
 Db 121 SMHAGTQGEKRSVEKYVFNQWELDINSRSLVSDGYSYKLPFRSEFRALLHFCENPGKI 180
 Qy 181 SRABELLKQMTGRELKPQDRTVDVTRIRKHFESTPDTPPEIATIHGEGYRFGCDLE 238
 Db 181 TRADLLKQMTGRELKPQDRTVDVTRIRKHFESVSGTPEIATIHGEGYRFGCDLE 238

RESULT 7
 I64099
 aerobic respiration control protein arca homolog - Haemophilus influenzae (strain Rd KW2
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
 C:Accession: I64099
 R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
 Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: I64099
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-236 <TIGR>
 A:Cross-references: UNIPROT:P44918; UNIPARC:UPI0000125D8D; GB:U32770; GB:L42023; NID:915
 C:Superfamily: ompR protein; response regulator homology
 C:Keywords: DNA binding; phosphoprotein; transcription regulation
 P:6-114/Domain: response regulator homology <RRH>
 P:54/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 78.4%; Score 964.5; DB 2; Length 236;
 Best Local Similarity 76.2%; Pred. No. 8.8e-69;
 Matches 183; Conservative 26; Mismatches 24; Indels 7; Gaps 2;

Qy 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATGAEMHQILSYDINLVIMDINLPKG 60
 Db 1 MTPPKLVVEDEIVTRNTLKGIFEAEGYDVFEAENGEMHILANHLNVLVMDINLPKG 60

Qy 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Db 61 NGLLLARELREQANVALMFLTRGDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120

Qy 121 TMNLGTVSEERSV---ESYKFNQWELDINSRSLIGPDGEQYKLPFRSEFRAMLHFCENPG 177
 Db 121 AM----PHQEKENTFGREYFNGKGLDNLNLSLITPEGQBFKLPFRSEFRAMLHFCENPG 176

Qy 178 KIQSRAELLKQMTGRELKPQDRTVDVTRIRKHFESTPDTPPEIATIHGEGYRFGCDLE 237
 Db 177 KQTRBELLKQMTGRELKPQDRTVDVTRIRKHFEDHPNTPNIMTIHGEYRFGCDIE 236

RESULT 8
 G82164
 DNA-binding response regulator TorR VC1719 [imported] - Vibrio cholerae (strain N16961 s
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C:Accession: G82164
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.J.;
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: AB2035; MUID:20406833; PMID:10952301
 A:Accession: G82164
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-234 <HEI>
 A:Cross-references: UNIPROT:Q9KRC3; UNIPARC:UPI00000C30AD; GB:AE004250; GB:AE003852; NID
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1719

A:Map position: 1

C:Superfamily: ompR protein; response regulator homology

Query Match 43.5%; Score 535; DB 2; Length 234;
Best Local Similarity 44.1%; Pred. No. 6.1e-35;
Matches 104; Conservative 49; Mismatches 73; Indels 10; Gaps 3;

Qy 5 HILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVMDINLPKNGILL 64
Db HVLVVEDDLVTRKSLSGYFQNEGYQVTEAESGAQREALELNDIDLVDLIDNLPDNGILL 63

Qy 65 LARELREQANVALMFLTGRDNEVDKILGLETGADDDYITKPNPRELTIARNLISRTNWL 124
Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 120

Qy 125 GTVSEERSVESYKFNWELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKISR 179
Db SLAQREIRITDKHDDKLVHFGEMTFDQRRAL-SRNGEPVKLTAKAYELLVALSSYPNQV 178

Qy 180 QSRALLKMTGRELKPHDRITVDVTIRIRKHFSTPTDPELIATHGEGYRFGCD 235
Db LSRERILAMISHRVDPADRTIDVLIRMRKAMVMDPKNPQIFVTVHGEYMFAGD 234

RESULT 9

A:4841

torCAD operon transcription regulator torR - Escherichia coli (strain K-12)
N:Alternate names: TMAO response regulator; torCAD response regulator
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: A64841; 169716; S77583
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64841
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-230 <BLAT>
A:Cross-references: UNIPROT:P38684; UNIPARC:UPI0000137179; GB:AE000201; GB:U00096; NID:9
A:Experimental source: strain K-12, substrain MG1655
R:Simon, G.; Mejean, V.; Jourlin, C.; Chippaux, M.; Pascal, M.C.
J. Bacteriol. 176, 5601-5606, 1994
A:Title: The torR gene of Escherichia coli encodes a response regulator protein involved
A:Reference number: 155071; MUID:94364937; PMID:8083154
A:Accession: 169716
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-133, 'L', 135-230 <RES>
A:Cross-references: UNIPARC:UPI000016F53A; EMBL:X78195; NID:g1122855; PIDN:CAAG3922.1; F
R:Jourlin, C.; Bengrine, A.; Chippaux, M.; Mejean, V.
Mol. Microbiol. 20, 1297-1306, 1996
A:Title: An unorthodox sensor protein (TorS) mediates the induction of the tor structural
A:Reference number: S77581; MUID:96405651; PMID:8809780
A:Accession: S77583
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133, 'L', 135-230 <JOU>
A:Cross-references: UNIPARC:UPI000016F53A; EMBL:X94231; NID:g1122855; PIDN:CAAG3922.1; F

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995

C:Genetics:

A:Gene: torR

C:Function:

A:Description: transcriptional activator for the torCAD operon

A>Note: member of a two-component regulatory system

C:Superfamily: ompR protein; response regulator homology

C:Keywords: DNA binding; phosphoprotein; transcription regulation

F:5-113/Domain: response regulator homology <RRH>

F:53/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match

Best Local Similarity

Matches 99; Conservative

38.9%; Score 479; DB 2; Length 230;

42.5%; Pred. No. 1.5e-30;

Mismatches 51; Indels 8; Gaps 4;

Qy 5 HILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVMDINLPKNGILL 64

Db HVLVVEDDLVTRKSLSGYFQNEGYQVTEAESGAQREALELNDIDLVDLIDNLPDNGILL 63

Qy 65 LARELREQANVALMFLTGRDNEVDKILGLETGADDDYITKPNPRELTIARNLISRTNWL 124

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 122

Qy 125 GTVSEERSVESYKFNWELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKISR 184

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 181

Qy 123 ARQAQPYTQDNCYRFPAGYCLNV-SRHTLERDGEPIKLTAEYEMLVAFVTPNGEILSR 181

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 181

Qy 185 LLKQMTGRELK-PHRTVDVTIRIRKHFSTPTDPELIATHGEGYRFGCDL 236

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 229

Qy 182 LLRLSARRVENPDRLTVDLIRLRHKLSA-----DLLVTQHGEGYFLAADV 229

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 229

RESULT 10

B85635

response transcription regulator for torA [imported] - Escherichia coli (strain O157:H7;

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004

C:Accession: B85635

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: B85635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <STO>

A:Cross-references: UNIPROT:P58357; UNIPARC:UPI0000137178; GB:AE005174; NID:g12514262; J

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: torR

C:Superfamily: ompR protein; response regulator homology

Query Match 38.7%; Score 476; DB 2; Length 230;

Best Local Similarity 42.5%; Pred. No. 2.7e-30;

Matches 99; Conservative 51; Mismatches 75; Indels 8; Gaps 4;

Qy 5 HILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVMDINLPKNGILL 64

Db HVLVVEDDLVTRKSLSGYFQNEGYQVTEAESGAQREALELNDIDLVDLIDNLPDNGILL 63

Qy 65 LARELREQANVALMFLTGRDNEVDKILGLETGADDDYITKPNPRELTIARNLISRTNWL 124

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 122

Qy 125 GTVSEERSVESYKFNWELDINSRLIGPDGEQYKLPFRSEFRAMLHFCENPGKISR 184

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 181

Qy 123 ARQAQPYTQDNCYRFPAGYCLNV-SRHTLERDGEPIKLTAEYEMLVAFVTPNGEILSR 181

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 181

Qy 185 LLKQMTGRELK-PHRTVDVTIRIRKHFSTPTDPELIATHGEGYRFGCDL 236

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 229

Qy 182 LLRLSARRVENPDRLTVDLIRLRHKLSA-----DLLVTQHGEGYFLAADV 229

Db LTRALREKSTVGIIILVTGRSDRIDRVGLEMGADDDYTKPLELRELVVVRKNLLWR---- 229

RESULT 11

F90772

response transcription regulator for torA [imported] - Escherichia coli (strain O157:H7;

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: F90772

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: F90772

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <HAY>

A:Cross-references: UNIPROT:P58357; UNIPARC:UPI0000137178; GB:BA000007; PIDN:BAB34573.1;

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1150

C:Superfamily: ompR protein; response regulator homology

Query Match 38.7%; Score 476; DB 2; Length 230;

Best Local Similarity 42.5%; Pred. No. 2.7e-30;

Matches 99; Conservative 51; Mismatches 75; Indels 8; Gaps 4;

Qy 5 HILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 64

Db 4 HIVIVEDPVTQARLOSFTQGYTVSVTASGAREIMQVQVDVLLDINLPDENGIM 63

Qy 65 LARELRQANVALMPLTGRDNEVDKILGLETGADDDYITKPFNPRLTIRARNLLSRWNL 124

Db 64 LTRALRERSTVGILLVTRGSDRIDRVGLEMGADDDYVTKPLELRELVVVRKNLLWR-IDL 122

Qy 125 GTVSEERSVSYKFNKGWELDINSRLIGPDGEQYKLPKPRSEFRAMLHFCENPGKIQSRAE 184

Db 123 ARQAQPYTQDNCYPAGYCLNV-SRHTLERDGEPIKLTAEYEMLVAFVTPNPGELSRER 181

Qy 185 LLKXMTGRELK-PHRTDVTIRIRKHPESFPDPEIATIHGEGYRFGDGL 236

Db 182 LLRLSARVENPDURTVDVLRRLRHKLSA-----DLVTQHGEGYFLAADV 229

RESULT 12

H83022

probable two-component response regulator PA4983 [imported] - Pseudomonas aeruginosa (st

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: H83022

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83022

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-244 <STO>

A:Cross-references: UNIPROT:Q9HU12; UNIPARC:UPI00000C5EAC; GB:AE004911; GB:AE004091; NID

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4983

C:Superfamily: ompR protein; response regulator homology

Query Match 37.2%; Score 458; DB 2; Length 244;

Best Local Similarity 39.4%; Pred. No. 7.6e-29;

Matches 95; Conservative 47; Mismatches 91; Indels 8; Gaps 3;

Qy 4 PHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSE-----YDINLVIMDINLP 58

Db 5 PRVLVVDVDPVIRELLQAYLGESEGYDVLCAAGNAEQAEACALHAGQFVELVLDIRLP 64

Qy 59 GNGGLLARELRQANVALMPLTGRDNEVDKILGLETGADDDYITKPFNPRLTIRARNLL 118

Db 65 GNDGLTLRELKVRSEVGIILITGRNDEIDRVLGECGADDDYVTKPLNPRELVSRKNLI 124

Qy 119 SRTMNL-GTVSEERSVSYKFNKGWELDINSRLIGPDGEQYKLPKPRSEFRAMLHFCENPG 177

Db 125 RRVHAQASAGPARQALR-QFGDWLLDADRRLIDHAGNETLLTHGSEFQLLAGFLRNSG 182

Qy 178 KIOSRAELLKXMTGRELKPHDRTVDVTIRIRKHPESFPDPEIATIHGEGYRFGDGL 237

Db 183 HTLSRDQLMDQIRNREWLPSRSIDVLVGLRRLKRLDPAPBQLIITIHGAGYLTAAAS 242

Qy 238 D 238

Db 243 D 243

RESULT 13

AF3608

two component response regulator [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004

C:Accession: AF3608

R:DelVecchio, V.G.; Kapratl, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,

.; Mazur, M.; Goltzman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens

A:Reference number: AD3252; PMID:11756688

A:Accession: AF3608

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-238 <KUR>

A:Cross-references: UNIPROT:Q8YBU5; UNIPARC:UPI0000058604; GB:AE008918;

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME110791

A:Map position: II

C:Superfamily: ompR protein; response regulator homology

Query Match 34.7%; Score 427; DB 2; Length 238;

Best Local Similarity 40.3%; Pred. No. 2e-26;

Matches 94; Conservative 44; Mismatches 93; Indels 2; Gaps 1;

Qy 5 HILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 64

Db 6 HILIVDDKDIRDLLHRELFKRGHVSITACNGDEMLVLSRTPDLDVLDVLMFKSGIE 65

Qy 65 LARELRQANVALMPLTGRDNEVDKILGLETGADDDYITKPFNPRLTIRARNLLSRWNL 124

Db 66 ICQDVRRTSRVPIIMLTATADAAADKILGLETGADDDYITKPFNPRLTIRARNLLSRWNL 123

Qy 125 GTVSEERSVSYKFNKGWELDINSRLIGPDGEQYKLPKPRSEFRAMLHFCENPGKIQSRAE 184

Db 124 GNRSPQALTAQIRFAGTCLDCARRLTSPHDVRELVTAEFNLLEAFVKSSQHLTSRQ 183

Qy 185 LLKXMTGRELKPHDRTVDVTIRIRKHPESFPDPEIATIHGEGYRFGDGL 237

Db 184 LMENAGHQAQVGYDRSDVLDILSLRKLKLEDDPCAPKLILITIRGGYQFGHEVE 236

RESULT 14

S28674

hypothetical protein 2 - Rhizobium sp. (strain IC3342)

C:Species: Rhizobium sp.

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: S28674

R:Upadhyaya, N.M.; Scott, K.F.; Tucker, W.T.; Watson, J.M.; Dart, P.J.

Mol. Plant Microbe Interact. 5, 129-143, 1992

A:Title: Isolation and characterization of Rhizobium (IC3342) genes that determine leaf

A:Reference number: S28673; MUID:92314428; PMID:1319772

A:Accession: S28674

A:Molecule type: DNA

A:Residues: 1-238 <UPA>

A:Cross-references: UNIPROT:Q53167; UNIPARC:UPI00000B7CA9; EMBL:M38698; NID:gl52257; PID

C:Superfamily: ompR protein; response regulator homology

C:Keywords: phosphoprotein

F:4-111/Domain: response regulator homology <RRH>

F:51/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 34.1%; Score 420; DB 1; Length 238;

Best Local Similarity 38.4%; Pred. No. 7.2e-26;

Matches 89; Conservative 49; Mismatches 86; Indels 6; Gaps 3;

Qy 4 PHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 63


```

Db      2 PHLIIVDDDPRIIRSMISRYLEDEGPRVLAENISQLREVLP-SVDLVLLDGLPGNGL 60
QY      64 LLARELREQANVALMFLTGRNEVDKILGLEIGADDYITKFPNPRELTIRARNLL---SR 120
Db      61 ELAREIDANFRVPTTIIVSGRDDVDRIIIGLEMGADDYVSKPFLNRELLARVSRVLR 120
QY      121 TNLGTVSEERSVESYKFNWELDINSRLIGPDGEQYKLPKRSFRAMLHFCENPGKIQ 180
Db      121 AMPLGPAQKATAGI--PHFDGMLDADRQLTSTAGTIETTGTFDLMVFTVTHFRVL 178
QY      181 SRAELLKKMTGRELKPHDRTVDVTIRIRKHFEPTDPTPEIATIHGEGYRF 232
Db      179 TRDFLLDQTRGTREAFDRAIDVQVTRLRAKVEDDGDPRRIKSVRGAGYVF 230

RESULT 15
F83133
Probable two-component response regulator PA4101 [imported] - Pseudomonas aeruginosa (st
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: F83133
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <STO>
A:Cross-references: UNIPROT:Q9HWS7; UNIPARC:UPI00000C5COF; GB:AE004826; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4101
C:Superfamily: ompR protein; response regulator homology

Query Match      32.5%; Score 400; DB 2; Length 246;
Best Local Similarity 38.2%; Pred. No. 2.8e-24;
Matches 89; Conservative 44; Mismatches 92; Indels 8; Gaps 3;

QY      5 HLIIVDDDELVTNTLKSIFAEQYDVFEATDGAEMHQILSEYDINLVIMDINLPGKNGLL 64
Db      6 HLIIVDDDELVTNTLKSIFAEQYDVFEATDGAEMHQILSEYDINLVIMDINLPGKNGLL 65
QY      65 LARELREQANVA--LMFLTGRNEVDKILGLEIGADDYITKFPNPRELTIRARNLLSRM 122
Db      66 LCRELVRGKHKATPVMLTARNDETDRIIGLEMGADDYLTTPFSARELLARINAVLRTR 125
QY      123 NLG---TVSEERSVESYKFNWELDINSRLIGPDGEQYKLPKRSFRAMLHFCENPGKI 179
Db      126 MLPPLNLTVSESRLLI---GFGQWQDTSARHLDDAGTVALSGAEYRLRLRVFLDHPORV 182
QY      180 QSRABELLKKMTGRELKPHDRTVDVTIRIRKHFEPTDPTPEIATIHGEGYRF 232
Db      183 LSRDQLNLNTQCREADIFDRSIDLLVSLRQLRGDDAREPEYIKTVSRSEGYVF 235
```

Search completed: February 11, 2006, 02:43:26
Job time : 45 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:35:03 ; Search time 254 Seconds
(without alignments)
661.086 Million cell updates/sec

Title: US-10-613-990A-32
Perfect score: 1230
Sequence: 1 MTPHILIVEDLVTRNLK.....PELIATHGEGYRFGDLED 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1230	100.0	238	1 ARCA_ECO57	P0a9q3 escherichia
2	1230	100.0	238	1 ARCA_ECOL6	P0a9q2 escherichia
3	1230	100.0	238	1 ARCA_ECOL1	P0a9q1 escherichia
4	1230	100.0	238	1 ARCA_SHIFL	P0a9q4 shigella fl
5	1227	99.8	238	2 Q57G13_SALCH	Q57g13 salmonella
6	1227	99.8	238	2 Q7CP63_SALTY	Q7cp63 salmonella
7	1227	99.8	238	2 Q8XFN2_SALTI	Q8xfn2 salmonella
8	1219	99.1	238	2 Q5PN99_SALPA	Q5pn99 salmonella
9	1145	93.1	238	2 Q66EU0_YERPS	Q66eu0 yersinia ps
10	1145	93.1	238	2 Q8ZIN7_YERPE	Q8zin7 yersinia pe
11	1144	93.0	238	2 Q6D0A5_ERWCT	Q6d0a6 erwinia car
12	1138	92.5	238	2 Q7N8Z7_PHOLL	Q7n8z7 photorhabdu
13	1098	89.3	238	2 Q9KPG9_VIBCH	Q9kpg9 vibrio chol
14	1073	87.2	238	2 Q87SD5_VIBHPA	Q87sd5 vibrio para
15	1064	86.5	207	2 Q5MDG2_ECOLI	Q5mdg2 escherichia
16	1061	86.3	206	2 Q5MDG1_ECOLI	Q5mdg1 escherichia
17	1060	86.2	239	2 Q6LUQ5_PROPR	Q6luq5 photobacter
18	1056	85.9	238	2 Q8DEN9_VIBVU	Q8den9 vibrio vuln
19	1056	85.9	238	2 Q7MNR9_VIBVU	Q7mnr9 vibrio vuln
20	1056	85.9	239	2 Q5E2Y1_VIBFV	Q5e2y1 vibrio fisc
21	1052	85.5	238	2 Q5EB54_VIBVU	Q5eb54 vibrio vuln
22	1051	85.4	204	2 Q5MDP8_ECOLI	Q5mdf8 escherichia
23	1043	84.8	203	2 Q5MDF3_ECOLI	Q5mdf3 escherichia
24	1012.5	82.3	236	2 Q9CP42_PASMO	Q9cp42 pasteurella
25	1006	81.8	238	2 Q8EAB9_SHEON	Q8eab9 shewanella
26	987.5	80.3	236	2 Q65SE9_MANSN	Q65se9 manheimia
27	972.5	79.1	236	2 Q4QM24_HAB18	Q4qm24 haemophilus
28	968	78.7	188	2 Q9K366_ECOLI	Q9k366 escherichia
29	968	78.7	237	2 Q7VP30_HABDU	Q7vp30 haemophilus
30	964.5	78.4	236	1 ARCA_HAEIN	P44918 haemophilus
31	963	78.3	188	2 Q9KH76_ECOLI	Q9kh76 escherichia

32	941	76.5	183	2	Q5IP84_ECOLI	Q5ip84 escherichia
33	941	76.5	183	2	Q5IP89_SHIFL	Q5ip89 shigella fl
34	941	76.5	183	2	Q5IP91_SHIDY	Q5ip91 shigella dy
35	941	76.5	183	2	Q5IP93_SHISO	Q5ip93 shigella so
36	936	76.1	182	2	Q5IP77_ECOLI	Q5ip77 escherichia
37	936	76.1	182	2	Q5IP98_SHISO	Q5ip98 shigella so
38	936	76.1	182	2	Q5IP99_SHIBO	Q5ip99 shigella bo
39	936	76.1	182	2	Q5IP95_SHIFL	Q5ip95 shigella fl
40	934	75.9	182	2	Q5IP88_SHIFL	Q5ip88 shigella fl
41	934	75.9	182	2	Q5IP94_SHIDY	Q5ip94 shigella dy
42	934	75.9	182	2	Q5IP81_ECOLI	Q5ip81 escherichia
43	930	75.6	181	2	Q5IP80_ECOLI	Q5ip80 escherichia
44	929	75.5	181	2	Q5IP90_SHIBO	Q5ip90 shigella bo
45	929	75.5	181	2	Q65C04_ECOLI	Q65c04 escherichia

ALIGNMENTS

RESULT 1

ID	ARCA_ECO57	STANDARD;	PRT;	238 AA.
AC	P0A9Q3;	P03026;		
DT	21-JUL-1986	(Rel. 01, Created)		
DT	21-JUL-1986	(Rel. 01, Last sequence update)		
DT	13-SEP-2005	(Rel. 48, Last annotation update)		
DE	Aerobic respiration control protein arcA.			
GN	Name=arcA; OrderedLocusNames=z6004, EC85359;			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OX	Enterobacteriaceae; Escherichia.			
NCBI	TaxID=83334;			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;			
RC	MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,			
RA	Apodaca J., Anaatharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL	Nature 409:529-533(2001).			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=O157:H7 / Sakai / RMD 0509952 / EHEC;			
RC	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT	O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL	DNA Res. 8:11-22(2001).			
CC	-!- FUNCTION: Member of the two-component regulatory system arcB/arcA.			
CC	Represses a wide variety of aerobic enzymes under anaerobic			
CC	conditions. It also may be involved in the osmoregulation of			
CC	envelope proteins. When activated by arcB, it negatively regulates			
CC	the expression of genes of aerobic function. Activates the			
CC	transcription of the plfB operon by binding to its promoter (By			
CC	similarity).			
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-!- PTM: Phosphorylated by arcB (By similarity).			
CC	-!- SIMILARITY: Contains 1 response regulatory domain.			
CC	-----			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use as long as its content is in no way modified and this statement is not			
CC	removed.			
CC	-----			
CC	EMBL; AE005174; AAG59581.1; -; Genomic_DNA.			

```

DR EMBL; BA000007; BAB38782.1; -; Genomic_DNA.
DR PIR; A86140; A86140.
DR PIR; G91298; G91298.
DR HSSP; P13792; 1MVO.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00072; Response_reg; 1.
DR Pfam; PF00486; Trans_reg_C; 1.
DR ProDom; PD000039; Response_reg; 1.
DR ProDom; PD000329; Trans_reg_C; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW Activator; Complete proteome; DNA-binding; Phosphorylation; Repressor;
KW Sensory transduction; Transcription; Transcription regulation;
KW Two-component regulatory system.
FT DOMAIN 1 118 Response regulatory.
FT MOD RES 54 54 4-aspartylphosphate (By similarity).
SQ SEQUENCE 238 AA; 27292 MW; 9316CF4D8EABD8 CRC64;

Query Match 100.0%; Score 1230; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 3.6e-88;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOTPHILIVEDLVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MOTPHILIVEDLVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPNPRELTIARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPNPRELTIARNLLSR 120
QY 121 TMNLGTVEERRSVESYKFNWELNDSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180
DB 121 TMNLGTVEERRSVESYKFNWELNDSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPEIATIHGEGYFCGDL 238
DB 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPEIATIHGEGYFCGDL 238

RESULT 2
ARCA_ECOLI
ID ARCA_ECOLI STANDARD; PRT; 238 AA.
AC POA9Q2; P03026;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Aerobic respiration control protein arca (Dye resistance protein).
GN Name=arca; Synonyms=cpxC, dye, fexA, msp, seg, sifa;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -!- FUNCTION: Member of the two-component regulatory system arcaB/arca.
CC Represents a wide variety of aerobic enzymes under anaerobic
CC conditions. It also may be involved in the osmoregulation of
CC envelope proteins. When activated by arcaB, it negatively regulates
CC the expression of genes of aerobic function. Activates the
CC transcription of the p1fB operon by binding to its promoter (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- PTM: Phosphorylated by arcaB (By similarity).

```

```

CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC -!- CAUTION: There are two genes termed arca in strain O6 of E.coli,
CC one refers to an arginine deiminase and the other to a two-
CC component regulator.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AE016772; AAN83908.1; -; Genomic_DNA.
CC HSSP; P13792; 1MVO.
CC InterPro; IPR001789; Response_reg.
CC InterPro; IPR001867; Trans_reg_C.
CC Pfam; PF00072; Response_reg; 1.
CC Pfam; PF00486; Trans_reg_C; 1.
CC ProDom; PD000039; Response_reg; 1.
CC ProDom; PD000329; Trans_reg_C; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW Activator; Complete proteome; DNA-binding; Phosphorylation; Repressor;
KW Sensory transduction; Transcription; Transcription regulation;
KW Two-component regulatory system.
FT DOMAIN 1 118 Response regulatory.
FT MOD RES 54 54 4-aspartylphosphate (By similarity).
SQ SEQUENCE 238 AA; 27292 MW; 9316CF4D8EABD8 CRC64;

Query Match 100.0%; Score 1230; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 3.6e-88;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOTPHILIVEDLVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MOTPHILIVEDLVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPNPRELTIARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPNPRELTIARNLLSR 120
QY 121 TMNLGTVEERRSVESYKFNWELNDSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180
DB 121 TMNLGTVEERRSVESYKFNWELNDSRLSGPGEQYKLPSEFRAMLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPEIATIHGEGYFCGDL 238
DB 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPEIATIHGEGYFCGDL 238

RESULT 3
ARCA_ECOLI
ID ARCA_ECOLI STANDARD; PRT; 238 AA.
AC POA9Q1; P03026;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Aerobic respiration control protein arca (Dye resistance protein).
GN Name=arca; Synonyms=cpxC, dye, fexA, msp, seg, sifa;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=85157583; PubMed=2984198;
RA Drury L.S., Buxton R.S.;
RT "DNA sequence analysis of the dye gene of Escherichia coli reveals
RT amino acid homology between the dye and OmpR proteins.";
RL J. Biol. Chem. 260:4236-4242(1985).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K12 / MG1655;

```

RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes";
RL Nucleic Acids Res. 23:2105-2119(1995).
[3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12";
RL Science 277:1453-1474(1997).
[4]
RP NUCLEOTIDE SEQUENCE OF 1-23.
RC STRAIN=K12;
RA Park S.J., Gunsalus R.P.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
[5]
RP PROTEIN SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12";
RL Electrophoresis 18:1259-1313(1997).
[6]
RP PROTEIN SEQUENCE OF 40-45 AND 123-128, AND CHARACTERIZATION.
RX MEDLINE=96037802; PubMed=7565118;
RA Drapal N., Sawers G.;
RT "Purification of ArcA and analysis of its specific interaction with
RT the pfl promoter-regulatory region";
RL Mol. Microbiol. 16:597-607(1995).
[7]
CC -!- FUNCTION: Member of the two-component regulatory system arcB/arcA.
CC Represses a wide variety of aerobic enzymes under anaerobic
CC conditions. Controls the resistance of *E. coli* to dyes; required
CC for expression of the alkaline phosphatase and sex factor P genes;
CC It also may be involved in the osmoregulation of envelope
CC proteins. When activated by arcB, it negatively regulates the
CC expression of genes of aerobic function. Activates the
CC transcription of the pflB operon by binding to its promoter.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: Phosphorylated by arcB.
CC -!- SIMILARITY: Contains 1 response regulatory domain.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

DR EMBL; M10044; AAA23718.1; -; Genomic DNA.
DR EMBL; U14003; AAA97297.1; -; Genomic DNA.
DR EMBL; U00096; AAC77354.1; -; Genomic DNA.
DR EMBL; L34010; AAA23476.1; -; Genomic DNA.
DR PIR; A03561; JYECR.
DR HSP; F13792; IMVO.
DR SWISS-2DPAGE; P0A9Q1; COLI.
DR EcoBASE; EB0059; -.
DR EcoGene; EG10061; arcA.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF0001867; Trans_reg_C.
DR Pfam; PF00072; Response_reg_1.
DR Pfam; PF00486; Trans_reg_C_1.
DR ProDom; PD000039; Response_reg_1.
DR ProDom; PD000329; Trans_reg_C_1.
DR SMART; SM00448; REC_1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Activator; Complete proteome; Direct protein sequencing; DNA-binding;
KW Phosphorylation; Repressor; Sensory transduction; Transcription;

KW Transcription regulation; Two-component regulatory system.
FT DOMAIN 1 118 Response regulatory.
FT MOD_RES 54 54 4-aspartylphosphate (By similarity).
SQ SEQUENCE 238 AA; 27292 MW; 9316CF4DE8EABDE8 CRC64;
Query Match 100.0%; Score 1230; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 3.6e-88;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQTPHILIVDELVTRNTLSKIFEAEGYDYFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MQTPHILIVDELVTRNTLSKIFEAEGYDYFEATDGAEMHQILSEYDINLVIMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKFPNPRELTIRARNLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKFPNPRELTIRARNLSR 120
QY 121 TMNLGTVSEERRSVESYKFNQWELDINSRLIGPDGEQYKLPSEPRAMHLFCENPGKI 180
DB 121 TMNLGTVSEERRSVESYKFNQWELDINSRLIGPDGEQYKLPSEPRAMHLFCENPGKI 180
QY 181 SRAELKKMTGRELKPDRVTVDVTRIRKHFSTPTDPTPELIATIHGEGYRFGCDLED 238
DB 181 SRAELKKMTGRELKPDRVTVDVTRIRKHFSTPTDPTPELIATIHGEGYRFGCDLED 238
RESULT 4
ID_ARCA_SHIFL STANDARD; PRT; 238 AA.
AC P0A9Q1; F03026;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DB 13-SEP-2005 (Rel. 48, Last annotation update)
DB Aerobic respiration control protein arcA.
GN Name=arcA; OrderedLocusNames=SF4433, S4704;
OS *Shigella flexneri*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
NCBI_TaxID=623;
[1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang J., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
RT through comparison with genomes of *Escherichia coli* K12 and O157";
RL Nucleic Acids Res. 30:4432-4441(2002).
[2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=2259274; PubMed=12704152;
RX DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of *Shigella*
RT flexneri serotype 2a strain 2457T";
RL Infect. Immun. 71:2775-2786(2003).
[3]
CC -!- FUNCTION: Member of the two-component regulatory system arcB/arcA.
CC Represents a wide variety of aerobic enzymes under anaerobic
CC conditions. It also may be involved in the osmoregulation of
CC envelope proteins. When activated by arcB, it negatively regulates
CC the expression of genes of aerobic function. Activates the
CC transcription of the pflB operon by binding to its promoter (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- PTM: Phosphorylated by arcB (By similarity).
CC -!- SIMILARITY: Contains 1 response regulatory domain.

```

CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AE005674; AAN45846.1; ALT INIT; Genomic_DNA.
CC EMBL; AE016993; AAP19620.1; -; Genomic_DNA.
CC HSP; P13792; IMVO.
CC InterPro; IPR001789; Response reg.
CC InterPro; IPR001867; Trans reg C.
CC Pfam; PF00072; Response reg; 1.
CC Pfam; PF00486; Trans_reg_C; 1.
CC ProDom; PD000039; Response_reg; 1.
CC ProDom; PD000329; Trans_reg_C; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS0110; RESPONSE REGULATORY; 1.
CC Activator; Complete proteome; DNA-binding; Phosphorylation; Repressor;
CC Sensory transduction; Transcription; Transcription regulation;
CC Two-component regulatory system.
CC DOMAIN 1 118 Response regulatory.
FT MOD RES 54 54 4-aspartylphosphate (By similarity).
SQ SEQUENCE 238 AA; 27292 MW; 9316CF4DE8EABD8 CRC64;

Query Match 100.0%; Score 1230; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 3.6e-88;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGVDVFEATDGAEMHQLSEYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGVDVFEATDGAEMHQLSEYDINLVIMDINLPKG 60

QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120

QY 121 TNNLGVTSSEERSVSYPKNGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
DB 121 TNNLGVTSSEERSVSYPKNGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKIQ 180

QY 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPTPEIIATIHGEGYRFGCDLED 238
DB 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPTPEIIATIHGEGYRFGCDLED 238

RESULT 5
ID Q57G13_SALCH PRELIMINARY; PRT; 238 AA.
AC Q57G13_SALCH
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Response regulator (OmpR family) in two-component regulatory system
DE with ArcB (Or CpxA), regulates genes in aerobic pathways.
GN Name-arcA; OrderedLocustNames=SC4443;
OS Salmonella choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE
RC STRAIN=SC-867;
RX PubMed=15781495; DOI=10.1093/nar/gki297;
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
RA Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
RT highly invasive and resistant zoonotic pathogen.";
RL Nucleic Acids Res. 33:1690-1698(2005).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC EMBL; AE017220; AAX68349.1; -; Genomic_DNA.
CC SMR; Q57G13; 1-122.
CC GO; GO:0003677; F:DNA binding; IEA.

```

```

DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0000150; P:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR InterPro; IPR011991; Wing_hlx_DNA_bd.
DR Pfam; PF00072; Response_reg; 1.
DR Pfam; PF00486; Trans_reg_C; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS0110; RESPONSE REGULATORY; 1.
KW Complete proteome; DNA-binding; Sensory transduction; Transcription;
KW Transcription regulation; Two-component regulatory system.
SQ SEQUENCE 238 AA; 27291 MW; 930B0F4DE8EABD8 CRC64;

Query Match 99.8%; Score 1227; DB 2; Length 238;
Best Local Similarity 99.6%; Pred. No. 6.2e-86;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGVDVFEATDGAEMHQLSEYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGVDVFEATDGAEMHQLSEYDINLVIMDINLPKG 60

QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120

QY 121 TNNLGVTSSEERSVSYPKNGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
DB 121 TNNLGVTSSEERSVSYPKNGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKIQ 180

QY 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPTPEIIATIHGEGYRFGCDLED 238
DB 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPTPEIIATIHGEGYRFGCDLED 238

RESULT 6
ID Q7CP63_SALTY PRELIMINARY; PRT; 238 AA.
AC Q7CP63_SALTY
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Response regulator (OmpR family) in two-component regulatory system
DE with ArcB (Or CpxA).
GN Name-arcA; OrderedLocustNames=STM4598;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP NUCLEOTIDE SEQUENCE
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC EMBL; AE008916; AAL23413.1; -; Genomic_DNA.
CC HSP; P08402; 1800.
CC SMR; Q7CP63; 1-122.
CC GO; GO:0003677; F:DNA binding; IEA.
CC GO; GO:0000156; P:two-component response regulator activity; IEA.
CC GO; GO:0007600; P:sensory perception; IEA.
CC GO; GO:0006350; P:transcription; IEA.
CC GO; GO:0000160; P:two-component signal transduction system (p. . .); IEA.
CC InterPro; IPR001789; Response_reg.
CC InterPro; IPR001867; Trans_reg_C.
CC Pfam; PF00072; Response_reg; 1.
CC Pfam; PF00486; Trans_reg_C; 1.

```

DR ProDom; PD000039; Response_reg; 1.
 DR SMART; PD000329; Trans_reg_C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS01110; RESPONSE REGULATORY; 1.
 KW Complete proteome; DNA-binding; Sensory transduction; Transcription;
 Transcription regulation; Two-component regulatory system.
 SQ SEQUENCE 238 AA; 27291 MW; 93080F4D8EABD8 CRC64;

Query Match 99.8%; Score 1227; DB 2; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.2e-88;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQTPHILVDELVTRNLTKSIFEAEGYDVPEATDGAEMHOILSEYDINLVIMDINLPKG 60
 Db 1 MQTPHILVDELVTRNLTKSIFEAEGYDVPEATDGAEMHOILSEYDINLVIMDINLPKG 60

Qy 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 120
 Db 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 120

Qy 121 TNLGTVSEERSVSEYKFNGLWELDINSRLIGPDGEQYKLPGRSEFRAMLHFCENPGKI 180
 Db 121 TNLGTVSEERSVSEYKFNGLWELDINSRLIGPDGEQYKLPGRSEFRAMLHFCENPGKI 180

Qy 181 SRAELLKMTGRELKHDPDRTVDVTRIRKHFESTPDPTEIATIHGSGYRFGDLED 238
 Db 181 SRAELLKMTGRELKHDPDRTVDVTRIRKHFESTPDPTEIATIHGSGYRFGDLED 238

RESULT 7
 Q8XFN2 SALT PRELIMINARY; PRT; 238 AA.
 AC Q8XFN2; Q7ALM2;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Global response regulator.
 GN Name=arcA; OrderedLocusNames=STY4947, t4637;
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RX DOI=10.1128/JB.185.7.2330-2337.2003;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 DR EMBL; AE016849; AAO72066.1; -; Genomic_DNA.
 DR EMBL; AL627284; CAD03429.1; -; Genomic_DNA.
 DR HSP; P08402; I800.
 DR SMR; Q8XFN2; 1-122.
 DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0000156; F:Two-component response regulator activity; IEA.
 DR GO; GO:0007600; P:sensory perception; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0000160; P:Two-component signal transduction system (p. . .); IEA.
 DR InterPro; IPR001789; Response_reg.
 DR InterPro; IPR001867; Trans_reg_C.
 DR InterPro; IPR011991; Wing_hlx_DNA_bd.
 DR Pfam; PF00072; Response_reg; 1.
 DR Pfam; PF00486; Trans_reg_C; 1.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD000329; Trans_reg_C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS01110; RESPONSE REGULATORY; 1.
 KW Complete proteome; DNA-binding; Sensory transduction; Transcription;
 Transcription regulation; Two-component regulatory system.
 SQ SEQUENCE 238 AA; 27291 MW; 93080F4D8EABD8 CRC64;

Query Match 99.8%; Score 1227; DB 2; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.2e-88;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQTPHILVDELVTRNLTKSIFEAEGYDVPEATDGAEMHOILSEYDINLVIMDINLPKG 60
 Db 1 MQTPHILVDELVTRNLTKSIFEAEGYDVPEATDGAEMHOILSEYDINLVIMDINLPKG 60

Qy 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 120
 Db 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 120

Qy 121 TNLGTVSEERSVSEYKFNGLWELDINSRLIGPDGEQYKLPGRSEFRAMLHFCENPGKI 180
 Db 121 TNLGTVSEERSVSEYKFNGLWELDINSRLIGPDGEQYKLPGRSEFRAMLHFCENPGKI 180

Qy 181 SRAELLKMTGRELKHDPDRTVDVTRIRKHFESTPDPTEIATIHGSGYRFGDLED 238
 Db 181 SRAELLKMTGRELKHDPDRTVDVTRIRKHFESTPDPTEIATIHGSGYRFGDLED 238

RESULT 8
 Q5PN99 SALPA PRELIMINARY; PRT; 238 AA.
 AC Q5PN99;
 DT 01-FEB-2005 (TReMBLrel. 29, Created)
 DT 01-FEB-2005 (TReMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TReMBLrel. 29, Last annotation update)
 DE Global response regulator.
 GN Name=arcA; OrderedLocusNames=SPA4408;
 OS Salmonella paratyphi-a.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=54388;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 9150 / SARB42;
 RX PubMed=15531882; DOI=10.1038/ng1470;
 RA McLeod M., Sanderson K.E., Clifton S.W., Latreille P.,
 RA Powollik S., Sabo A., Meyer R., Bieri T., Ozersky P., McLellan M.,
 RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
 RA Kohlberg S., Strong C., Du P., Carter J., Kremizki C., Layman D.,
 RA Leonard K., Sun H., Fulton L., Nash W., Miner T., Minx P., Florea L.,
 RA Delehaunt K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
 RA Spieth J., Wilson R.K.;
 RT "Comparison of genome degradation in Paratyphi A and Typhi, human-
 restricted serovars of Salmonella enterica that cause typhoid";
 RL Nat. Genet. 36:1268-1274(2004).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 DR EMBL; CP000026; AA080130.1; -; Genomic_DNA.
 DR SMR; Q5PN99; 1-122.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0000156; P:Two-component response regulator activity; IEA.
 DR GO; GO:0007600; P:sensory perception; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0000160; P:Two-component signal transduction system (p. . .); IEA.

DR InterPro; IPR001789; Response reg.
 DR InterPro; IPR001867; Trans_reg_C.
 DR Pfam; PF00072; Response_reg; 1.
 DR Pfam; PF00486; Trans_reg_C; 1.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD000329; Trans_reg_C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
 KW Complete proteome; DNA-binding; Sensory transduction; Transcription; Transcription regulation; Two-component regulatory system;
 SQ SEQUENCE 238 AA; 27261 MW; 9309DD9D8EABEEB CRC64;

Query Match 99.1%; Score 1219; DB 2; Length 238;
 Best Local Similarity 99.2%; Pred. No. 2.6e-87;
 Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MOTPHILVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
 Db 1 MOTPHILVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
 Qy 61 NGLLARELREQANVALMFLTRGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Db 61 NGLLARELREQANVALMFLTRGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Qy 121 TNNLGTVEERRSVESYKFNGLWELINSRLIGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
 Db 121 TNNLGTVEERRSVESYKFNGLWELINSRLIGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
 Qy 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPPEIIATIHGEGYRFGCDLED 238
 Db 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPPEIIATIHGEGYRFGCDLED 238

RESULT 9
 Q66EU0 YERPS PRELIMINARY; PRT; 238 AA.
 AC Q66EU0
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Response regulator (OmpR family), in two-component regulatory system
 DE with ArcB (Or CpxA), regulates genes in aerobic respiration.
 GN Name=arcA; OrderedLocNames=YPT0601;
 OS Versinia pseudotuberculosis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=633;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=IP32953 / Serotype I;
 RX PubMed=15358598; DOI=10.1073/pnas.0404012101;
 RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
 RA Regala W.M., Georgescu A.M., Verges L.M., Land M.L., Motin V.L.,
 RA Brubaker R.R., Fowler J., Himmelsbach J., Marceau M., Medigue C.,
 RA Simonet M., Chenal-Francoise V., Souza B., Dacheux D., Elliott J.M.,
 RA Derbise A., Hauser L.J., Garcia E.;
 RT "Insights into the evolution of Yersinia pestis through whole-genome
 comparison with Yersinia pseudotuberculosis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
 DR EMBL; BX936398; CAH19841.1; -; Genomic_DNA.
 DR SMR; Q66EU0; 1-122.
 DR GO; GO:0003677; F.DNA binding; IEA.
 DR GO; GO:0000156; F:two-component response regulator activity; IEA.
 DR GO; GO:0000160; P:two-component signal transduction system (p. . .); IEA.
 DR InterPro; IPR001789; Response_reg.
 DR InterPro; IPR001867; Trans_reg_C.
 DR Pfam; PF00072; Response_reg; 1.
 DR Pfam; PF00486; Trans_reg_C; 1.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD000329; Trans_reg_C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
 KW Complete proteome.

SQ SEQUENCE 238 AA; 27261 MW; BCA4DA76B61CA8C0 CRC64;

Query Match 93.1%; Score 1145; DB 2; Length 238;
 Best Local Similarity 92.9%; Pred. No. 1.6e-81;
 Matches 221; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MOTPHILVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
 Db 1 MOTPHILVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
 Qy 61 NGLLARELREQANVALMFLTRGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Db 61 NGLLARELREQANVALMFLTRGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSR 120
 Qy 121 TNNLGTVEERRSVESYKFNGLWELINSRLIGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
 Db 121 TNNLGTVEERRSVESYKFNGLWELINSRLIGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
 Qy 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPPEIIATIHGEGYRFGCDLED 238
 Db 181 SRAELLKMTGRELKPHDRTVDVTRIRKHFESTPDTPPEIIATIHGEGYRFGCDLED 238

RESULT 10
 Q8ZIN7 YERPE PRELIMINARY; PRT; 238 AA.
 AC Q8ZIN7; Q74Q02; Q7CG70; 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Aerobic respiration control protein (Negative response regulator of
 genes in aerobic pathways, sensors, ArcB and CpxA).
 DE Name=arcA; Synonyms=cpxC, dye, fexA, map, seg, sfrA;
 GN OrderedLocNames=Yp3725, YPO0458, Y3721;
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebaihia M., James K.D., Churcher C.M., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RX DOI=10.1128/JB.184.16.4601-4611.2002;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=91001;
 RX PubMed=15368993;
 RA Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,
 RA Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,
 RA Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,
 RA Yang H., Wang J., Huang P., Yang R.;
 RT "Complete genome sequence of Yersinia pestis strain 91001, an isolate
 avirulent to humans.";


```

DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR InterPro; IPR011951; Wing_hlx_DNA_bd.
DR Pfam; PF00072; Response_reg; 1.
DR Pfam; PF00486; Trans_reg_C; 1.
DR ProDom; PD000039; Response_reg; 1.
DR ProDom; PD000329; Trans_reg_C; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Complete proteome; DNA-binding; Sensory transduction; Transcription;
KW Transcription regulation; Two-component regulatory system.
SQ SEQUENCE 238 AA; 27236 MW; 46DF244FC0C21D5D CRC64;

Query Match 92.5%; Score 1138; DB 2; Length 238;
Best Local Similarity 93.2%; Pred. No. 5.8e-81;
Matches 221; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60

QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPFPNPRELTIRARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPFPNPRELTIRARNLLSR 120

QY 121 TNNLGTVSEERSVSEYKFNWELDINSRLSPAGEYKLPKPRSEFRAMLHFCENPGKIQ 180
DB 121 TNNLGTVSEERSVSEYKFNWELDINSRLSPAGEYKLPKPRSEFRAMLHFCENPGKIQ 180

QY 181 SRAELLKMTGRELKPHDRVTVDVTRIRKHFESTPTDPTPEIIATIHGEGYFCGDL 237
DB 181 TRADLLKMTGRELKPHDRVTVDVTRIRKHFESTPTDPTPEIIATIHGEGYFCGDL 237

RESULT 13
Q9KJP9 VIBCH
ID Q9KJP9 VIBCH PRELIMINARY; PRT; 238 AA.
AC Q9KJP9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aerobic respiration control protein FexA.
DE OrderedLocusNames=VC2368;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,
RA Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483 (2000).
RL EMBL; AE004307; RA095511.1; -; Genomic_DNA.
DR FIR; H82084; H82084.
DR HSSP; P08402; 1B00.
DR SMR; Q9KJP9; 1-122.
DR TIGR; VC2368; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0000156; P:two-component response regulator activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR InterPro; IPR011951; Wing_hlx_DNA_bd.
DR Pfam; PF00072; Response_reg; 1.
DR Pfam; PF00486; Trans_reg_C; 1.

```

```

DR ProDom; PD000039; Response_reg; 1.
DR ProDom; PD000329; Trans_reg_C; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Complete proteome.
SQ SEQUENCE 238 AA; 27119 MW; 6F59FBFD5C4C2540 CRC64;

Query Match 89.3%; Score 1098; DB 2; Length 238;
Best Local Similarity 88.2%; Pred. No. 7.8e-76;
Matches 210; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60

QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPFPNPRELTIRARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYTKPFPNPRELTIRARNLLSR 120

QY 121 TNNLGTVSEERSVSEYKFNWELDINSRLSPAGEYKLPKPRSEFRAMLHFCENPGKIQ 180
DB 121 TNNLGTVSEERSVSEYKFNWELDINSRLSPAGEYKLPKPRSEFRAMLHFCENPGKIQ 180

QY 181 SRAELLKMTGRELKPHDRVTVDVTRIRKHFESTPTDPTPEIIATIHGEGYFCGDL 238
DB 181 TRADLLKMTGRELKPHDRVTVDVTRIRKHFESTPTDPTPEIIATIHGEGYFCGDL 238

RESULT 14
Q87SD5 VIBPA
ID Q87SD5 VIBPA PRELIMINARY; PRT; 238 AA.
AC Q87SD5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Aerobic respiration control protein FexA.
DE OrderedLocusNames=VP0489;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae.";
RL Lancet 361:743-749 (2003).
RL EMBL; BA000031; BAC58752.1; -; Genomic_DNA.
DR HSSP; P08402; 1B00.
DR SMR; Q87SD5; 1-122.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0000156; P:two-component response regulator activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR InterPro; IPR011951; Wing_hlx_DNA_bd.
DR Pfam; PF00072; Response_reg; 1.
DR Pfam; PF00486; Trans_reg_C; 1.
DR ProDom; PD000039; Response_reg; 1.
DR ProDom; PD000329; Trans_reg_C; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Complete proteome.
SQ SEQUENCE 238 AA; 26936 MW; 14429758CC899F78 CRC64;

Query Match 87.2%; Score 1073; DB 2; Length 238;
Best Local Similarity 86.6%; Pred. No. 7.1e-76;
Matches 206; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

```

```
QY 1 MCTPHILIVEDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPK 60
Db 1 MCTPHILIVEDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPK 60
QY 61 NGLLIARELREQANVALMFLGRDNEVDKILGLEIGADDDYTKPNPRELTIRARNLLSR 120
Db 61 NGLLIARELREQANVALMFLGRDNEVDKILGLEIGADDDYTKPNPRELTIRARNLLSR 120
QY 121 TMTNLGTVSEERSVSEYKFNWELDINSRLSGEQYKLPSEFRAMLHFCENPGKIQ 180
Db 121 TMTNLGTVSEERSVSEYKFNWELDINSRLSGEQYKLPSEFRAMLHFCENPGKIQ 180
QY 121 SMTNVAQVEKRSVEKYFNGWLVLDINSRLSPAGDSYKLPSEFRALLHFCENPGKIQ 180
Db 121 SMTNVAQVEKRSVEKYFNGWLVLDINSRLSPAGDSYKLPSEFRALLHFCENPGKIQ 180
QY 181 SRAELLKMTGRELKPHDRTVDVTTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED 238
Db 181 SRAELLKMTGRELKPHDRTVDVTTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED 238
QY 181 TRADLLKMTGRELKPHDRTVDVTTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED 238
Db 181 TRADLLKMTGRELKPHDRTVDVTTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED 238
```

RESULT 15

```
Q5MDG2_ECOLI PRELIMINARY; PRT; 207 AA.
AC Q5MDG2;
DT 01-FEB-2005 (TremBLrel. 29, Created)
DT 01-FEB-2005 (TremBLrel. 29, Last sequence update)
DE 01-FEB-2005 (TremBLrel. 29, Last annotation update)
DE Aerobic respirator control protein A (Fragment).
GN Name=arCA;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxid=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=044;
RX PubMed=15814998; DOI=10.1128/JCM.43.4.1776-1781.2005;
RA Nemoy L.L., Kotetishvili M., Tigmo J., Keefer-Norris A., Harris A.D.,
RA Perencevich E.N., Johnson J.A., Torpey D., Sulakvelidze A.,
RA Morris J.G. Jr., Stine O.C.;
RT "Multilocus sequence typing versus pulsed-field gel electrophoresis
RT for characterization of extended-spectrum beta-lactamase-producing
RT Escherichia coli isolates."
RL J. Clin. Microbiol. 43:1776-1781(2005).
DR EMBL; AY832936; AAV91697.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0007600; P:sensory perception; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00072; Response_reg_1.
DR Pfam; PF00486; Trans_reg_C_1.
DR ProDom; PD000039; Response_reg_1.
DR ProDom; PD000329; Trans_reg_C_1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW DNA-binding; Sensory transduction; Transcription;
KW Transcription regulation.
FT NON_TER 1
FT NON_TER 207
SQ SEQUENCE 207 AA; 23745 MW; D821137D4D703F9D CRC64;
```

Query Match 86.5%; Score 1064; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 3e-75;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 20 KSIPEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLLALLARELREQANVALMP 79
Db 1 KSIPEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLLALLARELREQANVALMP 60
QY 80 LTGRDNEVDKILGLEIGADDDYTKPNPRELTIRARNLLSRMTNLGTVSEERSVSEYKF 139
Db 61 LTGRDNEVDKILGLEIGADDDYTKPNPRELTIRARNLLSRMTNLGTVSEERSVSEYKF 120
```

```
QY 140 NGWELDINSRLSGEQYKLPSEFRAMLHFCENPGKIQSRAELLKMTGRELKPHDR 199
Db 121 NGWELDINSRLSGEQYKLPSEFRAMLHFCENPGKIQSRAELLKMTGRELKPHDR 180
QY 200 TVDVTIRIRKHFESTPDTPEIIATIH 226
Db 181 TVDVTIRIRKHFESTPDTPEIIATIH 207
```

Search completed: February 11, 2006, 02:42:38

Job time : 256 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:42:54 ; Search time 48 Seconds
(without alignments)
409.934 Million cell updates/sec

Title: US-10-613-990A-32
Perfect score: 1230
Sequence: 1 MQTPHILIVEDLVTRNTLK.....PEIIATIHGEGYRFGCDLED 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgm2_6/ptodata/1/iaa/5-COMB.pep.*
- 2: /cgm2_6/ptodata/1/iaa/6-COMB.pep.*
- 3: /cgm2_6/ptodata/1/iaa/H-COMB.pep.*
- 4: /cgm2_6/ptodata/1/iaa/PCUTUS-COMB.pep.*
- 5: /cgm2_6/ptodata/1/iaa/RE-COMB.pep.*
- 6: /cgm2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1197	97.3	241	2	US-09-489-039A-8983 Sequence 8983, Ap
2	1133	92.1	241	2	US-09-543-681A-6205 Sequence 6205, Ap
3	458	37.2	278	2	US-09-252-991A-32988 Sequence 32988, A
4	400	32.5	721	2	US-09-252-991A-28293 Sequence 28293, A
5	384.5	31.3	249	2	US-09-543-681A-6721 Sequence 6721, Ap
6	376	30.6	246	2	US-09-107-532A-4135 Sequence 4135, Ap
7	375	30.5	233	2	US-09-082-077-4 Sequence 4, Appli
8	375	30.5	233	2	US-09-710-279-480 Sequence 480, App
9	375	30.5	233	2	US-09-925-637-46 Sequence 46, Appl
10	375	30.5	234	2	US-09-134-001C-3063 Sequence 3063, Ap
11	375	30.5	239	2	US-08-956-171E-5239 Sequence 5239, Ap
12	375	30.5	239	2	US-08-781-986A-5239 Sequence 5239, Ap
13	374	30.4	294	2	US-09-489-039A-9766 Sequence 9766, Ap
14	373.5	30.4	260	2	US-09-489-039A-10893 Sequence 10893, A
15	373.5	30.3	239	1	US-08-286-819A-44 Sequence 44, Appl
16	372.5	30.3	239	2	US-08-980-357-44 Sequence 44, Appl
17	372.5	30.3	239	2	US-09-357-375-44 Sequence 44, Appl
18	370	30.1	276	2	US-09-252-991A-21689 Sequence 21689, A
19	365	29.7	234	1	US-08-850-118-2 Sequence 2, Appli
20	365	29.7	234	1	US-09-008-253-2 Sequence 2, Appli
21	365	29.7	234	2	US-09-093-335-2 Sequence 2, Appli
22	365	29.7	234	2	US-09-583-110-4584 Sequence 4584, Ap
23	365	29.7	252	2	US-09-107-433-3288 Sequence 3288, Ap
24	362	28.4	246	2	US-09-134-001C-3175 Sequence 3175, Ap
25	349.5	28.4	265	2	US-09-328-352-6073 Sequence 6073, Ap
26	346.5	28.2	234	2	US-09-634-238-220 Sequence 220, App
27	345.5	28.1	236	2	US-09-543-681A-8023 Sequence 8023, Ap

ALIGNMENTS

RESULT 1

US-09-489-039A-8983
; Sequence 8983, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIORITY FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8983
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-8983

Query Match	97.3%	Score 1197;	DB 2;	Length 241;
Best Local Similarity	97.1%	Pred. No. 3.2e-125;		
Matches 231;	Conservative 5;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MQTPHILIVEDLVTRNTLKSI	FEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP	CGK 60
Db	4	MQTPHILIVEDLVTRNTLKSI	FEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP	CGK 63
Qy	61	NGLLARELREQANVALMFLTG	RNEVDKILGLEIGADDYITKPNPRELTIARNLSR	120
Db	64	NGLLARELREQADVALMFLTG	RNEVDKILGLEIGADDYITKPNPRELTIARNLSR	123
Qy	121	TMAIIGTVESEERSVESYKFN	GWELDINSRSLIGPDGEQYKLP	PRSEFRAMLHFCENPGKIQ 180
Db	124	TMAIIGTVESEERSVESYKFN	GWELDINSRSLVSPNGEQYKLP	PRSEFRAMLHFCENPGKIQ 183
Qy	181	SRAELLKQMTGRELKPHDRT	VDVTIRIRKHFESTPDTPPEIIATIHGSGYRFGCDLED	238
Db	184	SRAELLKQMTGRELKPHDRT	VDVTIRIRKHFESTPDTPPEIIATIHGSGYRFGCDLQE	241

RESULT 2

US-09-543-681A-6205
; Sequence 6205, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 5239:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5239:
US-08-956-171E-5239

Query Match 30.5%; Score 375; DB 2; Length 239;
Best Local Similarity 35.3%; Pred. No. 1.9e-33;
Matches 84; Conservative 50; Mismatches 76; Indels 28; Gaps 4;
QY 6 ILIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 65
DB 11 VVVVDEKPIADILFNLKKEGYDVYCAVDGNDVLDIYEEDPVLDDIIMPGRDGMV 70
QY 66 ARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR----- 120
DB 71 CREVKKYEMPILMTAKDSEIDKVLGLELGGADYVTKPFSTRELIARVKANLRHHYQ 130
QY 121 TNLGTVSEERSVSEYKFNKGWELDINSRL-IGPD-----GEQYKLPSEFRAMLHF 172
DB 131 AODTGNVNE-----ITIKDIVPDAYSIKKRGSDIELTHREFELFHYL 175
QY 173 CENPKIQSRAELKKMTGRELKPHDRVTVDVTTIRIRKHFSTPTDPEIIATIHGEY 230
DB 176 SKHMGQVMTREHLLQTVMGYDFGVDVTRVDTIRRLREKIEDDPSPHPEYIVTRRGVY 233

RESULT 12
US-08-781-986A-5239
Sequence 5239, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5239:
SEQUENCE CHARACTERISTICS:
LENGTH: 239 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-781-986A-5239

Query Match 30.5%; Score 375; DB 2; Length 239;
Best Local Similarity 35.3%; Pred. No. 1.9e-33;
Matches 84; Conservative 50; Mismatches 76; Indels 28; Gaps 4;
QY 6 ILIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 65
DB 11 VVVVDEKPIADILFNLKKEGYDVYCAVDGNDVLDIYEEDPVLDDIIMPGRDGMV 70
QY 66 ARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR----- 120
DB 71 CREVKKYEMPILMTAKDSEIDKVLGLELGGADYVTKPFSTRELIARVKANLRHHYQ 130
QY 121 TNLGTVSEERSVSEYKFNKGWELDINSRL-IGPD-----GEQYKLPSEFRAMLHF 172
DB 131 AODTGNVNE-----ITIKDIVPDAYSIKKRGSDIELTHREFELFHYL 175
QY 173 CENPKIQSRAELKKMTGRELKPHDRVTVDVTTIRIRKHFSTPTDPEIIATIHGEY 230
DB 176 SKHMGQVMTREHLLQTVMGYDFGVDVTRVDTIRRLREKIEDDPSPHPEYIVTRRGVY 233

RESULT 13
US-09-489-039A-9766
Sequence 9766, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9766
LENGTH: 294
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9766

Query Match 30.4%; Score 374; DB 2; Length 294;
Best Local Similarity 34.6%; Pred. No. 3.3e-33;
Matches 80; Conservative 50; Mismatches 93; Indels 8; Gaps 3;
QY 5 HILIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 64
DB 55 HILVVDDDRDIRELIVDYLEKSGYRASGAANGKAMWSVLKNHQIDLIIVLDIMMPGEDGLT 114
QY 65 LARELR--EQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 122
DB 115 LCRQLRANPOODIPVLMLTARTDSDRLGLEMGADDYLIKPFVARELLARIKALRRTR 174

```
QY 123 NLG---TVSEERRSVESYKFNKGWELNINSLIGPDGEQYKLPSEFRAMLHFCENPGKI 179
Db 175 ALPPNLQTEAGRL----AFGWLMDTARHLDDSGAIVASGAYRLLRVFLDPORV 231
QY 180 OSRAELKKMTGRELKHDPDRDVTIRIRKHFFSTPTDPTPEIATIHGEY 230
Db 232 LNRDQLNLTQGRDAELPERSIDLVSRLRQLRREDAREPAVIKTVRSEGY 282

RESULT 14
US-09-489-039A-10893
; Sequence 10893, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10893
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10893

Query Match 30.4%; Score 373.5; DB 2; Length 260;
Best Local Similarity 36.5%; Pred. No. 3.1e-33;
Matches 85; Conservative 47; Mismatches 94; Indels 7; Gaps 4;

QY 6 ILIIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINIVMDINLPGKNGLL 65
Db 28 ILVDDDDMLRALLERYLTQGFQVRSVANAQMDRLLTRESFHLWLDLMLPGEDGLSI 87
QY 66 ARELREQAN-VALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSRWN- 123
Db 88 CRRLSQSNPMPITWVTAKGEEVDRIVGLGEGADDYIPKPNPRELLARIARVLRQANE 147
QY 124 -LGTVSEERRSVESYKFNKGWELNINSLIGPDGEQYKLPSEFRAMLHFCENPGKIQSR 182
Db 148 LPGAPSQEEAVTAFGKF---KLNLTGTFREDFRED-EPMLTSGEFAVLKALVSHPREPLSR 203
QY 183 AELLKMTGRELKHDPDRDVTIRIRKHFFSTPTDPTPEIATIHGEYRFGCD 235
Db 204 DKLMNLARGREYSAMERSIDVQISRLRRMVESDPAHPRYIQTVWGLGYVFPD 256

RESULT 15
US-08-286-819A-44
; Sequence 44, Application US/08286819A
; Patent No. 5871910
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYPTIDES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,819A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5871910man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
US-08-286-819A-44
```

```
Query Match 30.3%; Score 372.5; DB 1; Length 239;
Best Local Similarity 36.5%; Pred. No. 3.5e-33;
Matches 85; Conservative 47; Mismatches 94; Indels 7; Gaps 4;
```

```
QY 6 ILIIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINIVMDINLPGKNGLL 65
Db 7 ILVDDDDMLRALLERYLTQGFQVRSVANAQMDRLLTRESFHLWLDLMLPGEDGLSI 66
```

```
QY 66 ARELREQAN-VALMFLTGRDNEVDKILGLEIGADDYITKPNPRELTIRARNLLSRWN- 123
Db 67 CRRLSQSNPMPITWVTAKGEEVDRIVGLGEGADDYIPKPNPRELLARIARVLRQANE 126
```

```
QY 124 -LGTVSEERRSVESYKFNKGWELNINSLIGPDGEQYKLPSEFRAMLHFCENPGKIQSR 182
Db 127 LPGAPSQEEAVTAFGKF---KLNLTGTFREDFRED-EPMLTSGEFAVLKALVSHPREPLSR 182
```

```
QY 183 AELLKMTGRELKHDPDRDVTIRIRKHFFSTPTDPTPEIATIHGEYRFGCD 235
Db 183 DKLMNLARGREYSAMERSIDVQISRLRRMVESDPAHPRYIQTVWGLGYVFPD 235
```

```
Search completed: February 11, 2006, 02:44:20
Job time : 49 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:54:50 ; Search time 174 Seconds
(without alignments)
571.514 Million cell updates/sec

Title: US-10-613-990A-32
Perfect score: 1230
Sequence: 1 MQTPHILIVEDLVTRNTLK.....PEIIATIHGEGYRFGCDLED 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA Main:
- 1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1230	100.0	238	4	US-10-613-990A-32
2	1145	93.1	238	4	US-10-613-990A-20
3	964.5	78.4	236	3	US-09-815-242-11143
4	964.5	78.4	236	4	US-10-282-122A-58341
5	778	63.3	176	5	US-10-450-763-59111
6	400	32.5	236	4	US-10-282-122A-74434
7	389.5	31.7	237	4	US-10-282-122A-60754
8	386.5	31.4	235	4	US-10-282-122A-46382
9	386.5	31.4	235	4	US-10-282-122A-72172
10	376	30.6	233	3	US-09-815-242-4913
11	375	30.6	234	3	US-09-815-242-10748
12	376	30.6	234	4	US-10-282-122A-42560
13	376	30.6	238	4	US-10-282-122A-71628
14	375	30.6	261	4	US-10-282-122A-60754
15	375	30.5	233	3	US-09-815-242-5676
16	375	30.5	233	3	US-09-925-637-46
17	375	30.5	233	4	US-10-084-205-46
18	375	30.5	233	4	US-10-712-713-46
19	375	30.5	233	5	US-10-923-429-46
20	375	30.5	234	4	US-10-724-972A-6908
21	375	30.5	235	3	US-09-815-242-12674
22	375	30.5	235	3	US-09-815-242-13000
23	375	30.5	235	4	US-10-282-122A-44277
24	375	30.5	235	5	US-10-857-625-768
25	375	30.5	239	2	US-08-781-986A-5239
26	375	30.5	239	4	US-10-282-122A-70821
27	375	30.5	239	4	US-10-329-624-5239

28	372.5	30.3	239	5	US-10-952-915-44	Sequence 44, Appl
29	372	30.2	232	3	US-09-738-626-3954	Sequence 3954, Ap
30	372	30.2	232	6	US-11-006-098-136	Sequence 136, App
31	371	30.2	235	5	US-10-501-282-962	Sequence 962, App
32	365.5	29.7	244	4	US-10-282-122A-50681	Sequence 50681, A
33	365	29.7	226	4	US-10-282-122A-53860	Sequence 53860, A
34	365	29.7	234	3	US-09-815-242-13265	Sequence 13265, A
35	365	29.7	234	3	US-09-815-242-13558	Sequence 13558, A
36	365	29.7	234	5	US-10-472-928-2464	Sequence 2464, Ap
37	365	29.7	251	4	US-10-282-122A-74003	Sequence 74003, A
38	365	29.7	252	5	US-10-617-320-3288	Sequence 3288, Ap
39	362	29.4	246	4	US-10-724-972A-4146	Sequence 4146, Ap
40	361	29.3	230	5	US-10-501-282-960	Sequence 960, App
41	356	28.9	223	4	US-10-156-761-11506	Sequence 11506, A
42	356	28.9	232	4	US-10-282-122A-52443	Sequence 52443, A
43	352.5	28.7	227	4	US-10-282-122A-61935	Sequence 61935, A
44	352.5	28.7	227	4	US-10-282-122A-64389	Sequence 64389, A
45	352.5	28.7	230	4	US-10-282-122A-62590	Sequence 62590, A

ALIGNMENTS

RESULT 1

US-10-613-990A-32
; Sequence 32, Application US/10613990A
; Publication No. US20040180404A1
; GENERAL INFORMATION:
; APPLICANT: Ajinomoto Co., Inc.
; TITLE OF INVENTION: Method for Producing Target Substance by Fermentation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/613,990A
; PRIOR FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: JP 2002-203764
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 32
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-10-613-990A-32

Query Match	100.0%	Score 1230;	DB 4;	Length 238;
Best Local Similarity	100.0%	Pred. No. 4.2e-116;		
Matches	238;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
QY	1	MQTPHILIVEDLVTRNTLKSI	FEAEGYDVFEATDGAEMHQIL	SEYDINLVIMDINLPKG 60
DB	1	MQTPHILIVEDLVTRNTLKSI	FEAEGYDVFEATDGAEMHQIL	SEYDINLVIMDINLPKG 60
QY	61	NGLLARELREQANVALMFLTG	RNEVDKIIGLEIGADDYITKPN	PRELTIARNLISR 120
DB	61	NGLLARELREQANVALMFLTG	RNEVDKIIGLEIGADDYITKPN	PRELTIARNLISR 120
QY	121	TWNLTGVSEERSVESYKFN	GWELDIINSRSLIGPDGEQYKLP	RSSEFRAMLHFCENPGKIQ 180
DB	121	TWNLTGVSEERSVESYKFN	GWELDIINSRSLIGPDGEQYKLP	RSSEFRAMLHFCENPGKIQ 180
QY	181	SRABLLKMTQRELPKPHDRT	VDVTIRIRKHFESTPDTPPEI	ATIHGEGYRFGCDLED 238
DB	181	SRABLLKMTQRELPKPHDRT	VDVTIRIRKHFESTPDTPPEI	ATIHGEGYRFGCDLED 238

RESULT 2

US-10-613-990A-20
; Sequence 20, Application US/10613990A
; Publication No. US20040180404A1
; GENERAL INFORMATION:
; APPLICANT: Ajinomoto Co., Inc.
; TITLE OF INVENTION: Method for Producing Target Substance by Fermentation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/613,990A

```
; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: JP 2002-203764
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Pantoea ananatis
US-10-613-990A-20

Query Match          93.1%; Score 1145; DB 4; Length 238;
Best Local Similarity 92.0%; Pred. No. 1.7e-107;
Matches 219; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSYDINLVIMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIARNLLSR 120
QY 121 TNLGTVSEERSVSYKFNWELDINSRSLIGPDGEQYKLPKPRSEFRAMLHFCENPG 177
DB 121 TNLGTVSEERSVSYKFNWELDINSRSLIGPDGEQYKLPKPRSEFRAMLHFCENPG 177
QY 178 KIQRARELLKQMTGRELKHDPDRTVDVTRIRKHFESTPDPEIATIHGEGYRFGDLE 237
DB 178 KIQRARELLKQMTGRELKHDPDRTVDVTRIRKHFESTPDPEIATIHGEGYRFGDLE 237

RESULT 3
US-09-815-242-11143
; Sequence 11143, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Judith W.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11143
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11143

Query Match          78.4%; Score 964.5; DB 3; Length 236;
Best Local Similarity 76.2%; Pred. No. 3.5e-89;
Matches 183; Conservative 26; Mismatches 24; Indels 7; Gaps 2;

; CURRENT FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: JP 2002-203764
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Pantoea ananatis
US-10-613-990A-20

Query Match          93.1%; Score 1145; DB 4; Length 238;
Best Local Similarity 92.0%; Pred. No. 1.7e-107;
Matches 219; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSYDINLVIMDINLPKG 60
DB 1 MQTPHILIVEDLVTRNTLKSIFEAEGYDVFEATDGAEMHQILSYDINLVIMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIARNLLSR 120
DB 61 NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIARNLLSR 120
QY 121 TNLGTVSEERSVSYKFNWELDINSRSLIGPDGEQYKLPKPRSEFRAMLHFCENPG 177
DB 121 TNLGTVSEERSVSYKFNWELDINSRSLIGPDGEQYKLPKPRSEFRAMLHFCENPG 177
QY 178 KIQRARELLKQMTGRELKHDPDRTVDVTRIRKHFESTPDPEIATIHGEGYRFGDLE 237
DB 178 KIQRARELLKQMTGRELKHDPDRTVDVTRIRKHFESTPDPEIATIHGEGYRFGDLE 237

RESULT 4
US-10-282-122A-58341
; Sequence 58341, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58341
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-282-122A-58341

Query Match          78.4%; Score 964.5; DB 4; Length 236;
Best Local Similarity 76.2%; Pred. No. 3.5e-89;
Matches 183; Conservative 26; Mismatches 24; Indels 7; Gaps 2;
```

```
QY 1 MOTPHILIVEDLVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
Db 1 MTTPKILVVEDEIVTRNTLKGIFEAGYDVFEAENGVEWHILANHNINLVVMDINLPKG 60
QY 61 NGLLARELREQANVALMFLTRDNEVDKILGLEITGADDTYTKPNPRELTIARNLISR 120
Db 61 NGLLARELRELSLPLIFLTRDNEVDKILGLEITGADDTYTKPNPRELTIARNLISR 120
QY 121 TNLGTVSEERSV--ESYKFNWELDINSRLIGPDGQYKLPRESFRAMLHFCENPG 177
Db 121 AM----PQEKENTGRFYRNGWKLDSLSLTPGQEFKLPRESFRAMLHFCENPG 176
QY 178 KIQSRAELLKQMTGRELKHDPDRVDTVIRIRKHFESTPTPEIIATIHGEGYFCGDL 237
Db 177 KLOTREELLKQMTGRELKHDPDRVDTVIRIRKHFEDHPNPIIIMTIHGEGYFCGDIE 236

RESULT 5
US-10-450-763-59111
; Sequence 59111, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyscg, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59111
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (26)..(69)
; OTHER INFORMATION: Transcriptional regulatory protein C terminal, domain
; OTHER INFORMATION: identified by eMATRIX, accession number PF00486C, p-value=8.560e-
; OTHER INFORMATION: 34, raw score of 20.21
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (105)..(169)
; OTHER INFORMATION: Transcriptional regulatory protein, C ter domain identified
; OTHER INFORMATION: by Pfam, accession name trans_reg_C, E-value=2.4e-24, Pfam score
; OTHER INFORMATION: 85.8
US-10-450-763-59111

Query Match 63.3%; Score 778; DB 5; Length 176;
Best Local Similarity 93.8%; Pred. No. 1.9e-70;
Matches 152; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 70 REQANVALMFLTRDNEVDKILGLEITGADDTYTKPNPRELTIARNLISRMTNLGTVSE 129
Db 18 QKQANVALMFLTRDNEVDKILGLEITGADDTYTKPNPRELTIARNLISRMTNLGTVSE 77
QY 130 ERRSVESYKFNWELDINSRLIGPDGQYKLPRESFRAMLHFCENPGKIQSRAELLKQM 189
Db 78 ERRSVESYKFNWELDINSRLIGPDGQYKLPRESFRAMLHFCENPGKIQSRAELLKQM 137
QY 190 TGRELKHDPDRVDTVIRIRKHFESTPTPEIIATIHGEGYR 231
Db 138 TGRELKHDPDRVDTVIRIRKHFESTPTPEIIATIHGEGYR 175

RESULT 6
US-10-282-122A-74434
```

```
; Sequence 74434, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74434
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-282-122A-74434

Query Match 32.5%; Score 400; DB 4; Length 236;
Best Local Similarity 39.5%; Pred. No. 6.2e-32;
Matches 92; Conservative 45; Mismatches 82; Indels 14; Gaps 4;

QY 6 ILIVDELVTRNTLKSIFEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 65
Db 4 ILIVDEKPISDIIKFNLTKEGYDVTAFDGRAVTFEEKPDLIILDLMLPELDGLEV 63
QY 66 ARELREQANVALMFLTRDNEVDKILGLEITGADDTYTKPNPRELTIARNLISRMTNLG 125
Db 64 AKERKTSHPVPIIMLSAKOSFQKVI GLEIGADDTYTKPFNSRRELLARVKHLRRTETIE 123
QY 126 T-VSEERRSVSEYKFNWELDINSRLIGPD-----GEQYKLPRESFRAMLHFCENPG 177
Db 124 TAVAEANSSGTQ-----ELTIGNLIIL-PDAFVAKKHGQVELTHREFLLHLANENG 177
QY 178 KIQSRAELLKQMTGRELKHDPDRVDTVIRIRKHFESTPTPEIIATIHGEGY 230
Db 178 QVMTREHLLIEVMGYDVGDRVTDVTVRRLEKIEDTPSRPEVILTTRGVGY 230

RESULT 7
US-10-282-122A-60754
; Sequence 60754, Application US/10282122A
; Publication No. US20040029129A1
```

; GENERAL INFORMATION:

; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 60754

; LENGTH: 237

; TYPE: PRT

; ORGANISM: Listeria monocytogenes

US-10-282-122A-60754

Query Match 31.7%; Score 389.5; DB 4; Length 237;

Best Local Similarity 36.0%; Pred. No. 7.3e-31;

Matches 86; Conservative 53; Mismatches 81; Indels 19; Gaps 4;

QY 1 MCTPHILIVEDLVTRNTLKSIFBAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MAEKILVDDDKPIADILKFNLEKEGFEIWAHDGDEAIEKATEBEQDMLLDIMLPKG 60
QY 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDDYTKFPNPRELTIRARNLLSR 120
DB 61 DGLVCEVRKVKYDPIINVTAKDSEIDKVLGLEIGADDDYTKFPNPRELTIRARNLLSR 120
QY 121 --TNNLGTVSEERRSVESYKFNKGWELDINSRSLIGP-----GEQVKLPSEPRAM 171
DB 121 HSQVSSSTAEEBENS-----BLEIGS-LIHPDAYVASKRGTELTTHREPELLHY 170
QY 172 FCENFGKIQSRAELKKMTGRELKPHDRVTVDVTIRIRKHFESTDPTPEIATIHGEGY 230
DB 171 LAKHMGQVMTREHLLQTVGWGYDFGDVRTVDVTVRRLREKIEDNPNFPAVLVTRRGVY 229

RESULT 8

US-10-282-122A-46382

; Sequence 46382, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Lianguo

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 46382

; LENGTH: 235

; TYPE: PRT

; ORGANISM: Bacillus anthracis

US-10-282-122A-46382

Query Match 31.4%; Score 386.5; DB 4; Length 235;

Best Local Similarity 35.1%; Pred. No. 1.5e-30;

Matches 85; Conservative 46; Mismatches 84; Indels 27; Gaps 2;

QY 1 MCTPHILIVEDLVTRNTLKSIFBAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKG 60
DB 1 MMGKKILVDDDKPIADILKFNLEKEGFEIWAHDGDEAIEKATEBEQDMLLDIMLPKG 60
QY 61 NGLLARELRQANVALMFLTGRDNEVDKILGLEIGADDDYTKFPNPRELTIRARNLLSR 120
DB 61 DGLVCEVRKIRKSEMPIMLTAKDSEIDKVLGLEIGADDDYTKFPSTRELLARVANLRR 120
QY 121 TNNLGTVSEERRSVESYKFNKGWELDINSRSLIGP-----DGEQVKLPSEPR 168
DB 121 HQQGAABKEE-----NTEWIGPIVINPNAYSVTYKRENIELTHREPEL 165
QY 169 MLHFENFGKIQSRAELKKMTGRELKPHDRVTVDVTIRIRKHFESTDPTPEIATIHGE 228
DB 166 LHYLAKHLGQVMTREHLLQTVGWGYDFGDVRTVDVTVRRLREKIEDNFSHPFLIVTRGV 225
QY 229 QY 230
DB 226 QY 227

RESULT 9

US-10-282-122A-72172

; Sequence 72172, Application US/10282122A

; Publication No. US20040029129A1


```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10748
; LENGTH: 234
; TYPE: PRF
; ORGANISM: Enterococcus faecalis
; US-09-815-242-10748

Query Match          30.6%; Score 376;
Best Local Similarity 36.6%; Pred. No.
Matches 85; Conservative 49; Mismatch

Qy 6 ILIVEDELVTRNTUKSIFEAEAGYDVFEATD
Db 4 ILVVDDEKPISEIVKYNLVKESGEYVFTAYD

Qy 66 ARELREQANVALMPLTGRDNEVDKTLGLEI
Db 64 AREVRKTHDMPIIMVTKAKSDIKVLJGUELA

Qy 126 TVSBERRSEVSYKFNGLWEDINSRSLIGPD
Db 120 ATNAKEAVTTQS----ELTIGDLT-IHPDD

Qy 179 IQSRABLLKKWGTGRELKHPRDVTVDVIRRI
Db 175 VMTREHLLOTWGVYDFGQVTRVTDVTVRRLL

```

```

RESULT 12
US-10-282-122A-42560
; Sequence 42560, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

```

```
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42560
; LENGTH: 234
; TYPE: PR1
; ORGANISM: Enterococcus faecalis
US-10-282-122A-42560

Query Match          30.6%; Score 376; DB 4; Length 234;
Best Local Similarity   36.6%; Pred. No. 1.7e-29;
Matches    85; Conservative 49; Mismatches 82; Indels 16; Gaps 4;

Qy      6 ILIVEDLVTRNTLKSI FEAGYDVFEATDGAEMHQILSEYDINLVMIDNLPKNGLL 65
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Db      4 ILVVDDEKPISIVKYNLVKREVFTAYDGEAELEKVVEVPDLII L DMLPKMDGLEV 63
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :

Qy      66 ARELRQAENVALMFELICRDNEDVKIILEGCADDYTTPKNPFELITIRARNLLSRTNLG 125
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Db      64 AREVRKTHDPIMVTVAADSEIKDKVLGLEDGADDYVTKPPSNRELVARVKANLRR----G 119
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :

Qy      126 TVSERRSVESYKFNGWELDINSRLTGPD-----GEQYKLP RSEFRAMLHFCENPGK 178
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Db      120 ATNAKEAVTTQS-----ELTIGDLT-IHPDAYVMSKGEKIETHREFELYLYAKHGQ 174
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :

Qy      179 IQGRAELLKKMTGRELKHDPDRTDVDTIRRIKHFESTPDDPTPEIATHSGGY 230
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Db      175 VMTREHLDTFVGWDYDGVDGVDTVTVRUREKIEDSPSHPTVLTVRGGVY 226
        |||:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
```

```

RESULT 13
US-10-282-122A-71628
; Sequence 71628, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71628
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71628

Query Match      30.6%; Score 376; DB 4; Length 238;
Best Local Similarity 35.6%; Pred. No. 1.7e-29;
Matches 83; Conservative 52; Mismatches 80; Indels 18; Gaps 3;

QY 6 ILIVDELVTRNLKSIPEAEGYDVFEATDGAEMHQILSEYDINLVINDINLPKNGLL 65
Db 10 VVVVDEKPIADILEFNLUKKEGYDYVCAYDGNDAVDLIYESEPDVLLDMLPGRDGM 69
QY 66 ARELREQANVALMFLTGDRNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 125
Db 70 CHEVRKKEMPIMLTAKDSEIDKVLGLELGADDYITKPFNPRLTIRARNLLSR 124
QY 126 TVSEERRSVESYKFGNWELDINSRLTGPD-----GEQYKLPSPFRAMLHFC 177
Db 125 -----HYSQPAQEVSGTTNEITIKDIVIPDAYSIKRGEDIELTHRFELF 179
QY 178 KIQSRAELLKMTGRELPHDRTVDVTRIRKHPSTPDTEIATIHGEGY 230
Db 180 QVMTREHLLQTVMGYDFGVDVTRVTRIRKREKIEDDPSHPHYIVTRRGVGY 232

RESULT 14
US-10-282-122A-57574
; Sequence 57574, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57574
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Enterococcus faecium
; FEATURES:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)-(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (20)-(20)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-57574

Query Match      30.6%; Score 376; DB 4; Length 261;
Best Local Similarity 35.0%; Pred. No. 2e-29;
Matches 79; Conservative 55; Mismatches 88; Indels 4; Gaps 2;

QY 6 ILIVDELVTRNLKSIPEAEGYDVFEATDGAEMHQILSEYDINLVINDINLPKNGLL 65
Db 31 VLVVDEKPIADILEFNLUKKEGYDYVCAYDGNDAVDLIYESEPDVLLDMLPGRDGM 90
QY 66 ARELREQANVALMFLTGDRNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR 125
Db 91 AREVRKTYDMPIMLTAKDSEIDKVLGLELGADDYITKPFNPRLTIRARNLLSR 147
QY 126 TVSEERRSVESYKFGNWELDINSRLTGPD-----GEQYKLPSPFRAMLHFC 184
Db 148 TAAKEPEEAPAEALTIGDLTIHPEAYMVTKRGETIELTHRFELFYLAKHLGQVMTREH 207
QY 185 LLKMTGRELPHDRTVDVTRIRKHPSTPDTEIATIHGEGY 230
Db 208 LLQTVMGYDFGVDVTRVTRIRKREKIEDDPSHPHYIVTRRGVGY 253

RESULT 15
US-09-815-242-5676
; Sequence 5676, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 02:55:25 ; Search time 17 Seconds
(without alignments)
183.716 Million cell updates/sec

Title: US-10-613-990A-32
Perfect score: 1230
Sequence: 1 MPTPHILIVEDLVTRNTLK.....PEIIATHGEGYFCGBLED 238

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgm2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgm2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgm2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	375	30.5	233	6	US-10-793-626-480
2	285	23.2	238	7	US-11-074-176-148
3	274	22.3	225	6	US-10-467-657-2276
4	262	21.3	224	6	US-10-793-626-3266
5	204	16.6	166	6	US-10-793-626-1424
6	204	16.6	166	6	US-10-793-626-2788
7	198.5	16.1	185	6	US-10-873-528-54
8	175	14.2	122	7	US-11-098-686-11387
9	154	12.5	188	6	US-10-467-657-1982
10	146.5	11.9	425	6	US-10-467-657-8052
11	139	11.3	465	7	US-11-098-686-10932
12	131.5	10.7	141	6	US-10-793-626-686
13	131.5	10.7	141	6	US-10-793-626-2234
14	130.5	10.6	471	7	US-11-098-686-11324
15	128.5	10.4	457	7	US-11-098-686-11122
16	124	10.1	251	6	US-10-793-626-238
17	124	10.1	251	6	US-10-793-626-722
18	122	9.9	126	7	US-11-098-686-10771
19	121	9.8	218	6	US-10-467-657-4822
20	121	9.8	2087	7	US-11-075-185-28
21	114.5	9.3	461	7	US-11-075-185-25
22	114	9.3	255	7	US-11-075-185-20
23	113	9.2	316	7	US-11-098-686-10974
24	109.5	8.9	740	7	US-11-140-625-12
25	109	8.9	218	6	US-10-793-626-108

26	108	8.8	738	7	US-11-140-625-11	Sequence 11, Appl
27	105	8.5	132	6	US-10-667-295-147	Sequence 147, Appl
28	104	8.5	141	6	US-10-667-295-25	Sequence 25, Appl
29	104	8.5	166	6	US-10-667-295-24	Sequence 24, Appl
30	104	8.5	170	6	US-10-667-295-23	Sequence 23, Appl
31	96	7.8	754	7	US-11-140-625-10	Sequence 10, Appl
32	90.5	7.4	212	6	US-10-454-437-298	Sequence 298, Appl
33	88.5	7.2	319	7	US-11-098-686-11362	Sequence 11362, A
34	87.5	7.1	884	7	US-11-098-686-11269	Sequence 11269, A
35	85	6.9	142	6	US-10-667-295-74	Sequence 74, Appl
36	85	6.9	3056	7	US-11-109-156-20	Sequence 20, Appl
37	82	6.7	207	6	US-10-467-657-2592	Sequence 2592, Ap
38	82	6.7	210	6	US-10-454-437-302	Sequence 302, Appl
39	82	6.7	363	6	US-10-995-561-541	Sequence 541, Appl
40	82	6.7	725	6	US-10-467-657-1536	Sequence 1536, Ap
41	81.5	6.6	208	6	US-10-793-626-782	Sequence 782, Appl
42	81	6.6	131	6	US-10-667-295-34	Sequence 34, Appl
43	81	6.6	136	6	US-10-667-295-33	Sequence 33, Appl
44	81	6.6	341	6	US-10-515-481-3	Sequence 3, Appl
45	79.5	6.5	738	7	US-11-140-625-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-793-626-480
; Sequence 480, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793.626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 480
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-480

Query Match	30.5%	Score 375;	DB 6;	Length 233;
Best Local Similarity	35.6%	Pred. No. 1.5e-26;		
Matches	83;	Conservative 52;	Mismatches 80;	Indels 18; Gaps 3;
QY	6	ILIVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILISEYDINILVINDINLPGKNGLLL	65	
Db	5	VVVVDEKPIADILFNLKKGEDYVCAYDNGDAVDLIYESEPDIVLADIMLPGRDGMV	64	
QY	66	ARELREQANVAMFLTGRDNEVDKILGLEIGADDYITKFPNPRELTIRARULLSRMTNLG	125	
Db	65	CREVRKKYEMFIIMLTAKDSIDKVLGELGADDYVTKFPFTRELIARVKNLR	119	
QY	126	TVSEBRSVESYKFNWELDINSRL-IGPD-----GEQYKLPFRSEFRAMLHFCENPG	177	
Db	120	-----HYSQPAQVSGATNITIKDIVIYPDAYSIKKRGEDIETHREFELPHYLSKMG	174	
QY	178	KIQSRAELLKQMGRELKPHDRTVDIRRIRKHFESTPTDPEIATHHGEGY	230	
Db	175	QVMTREHLLQTVMGVDYFGDVRTVDVITIRLRREKIEDDPSPHEVIVTRRGVY	227	

RESULT 2
US-11-074-176-148
; Sequence 148, Application US/11074176
; Publication No. US20050250135A1

GENERAL INFORMATION:
; APPLICANT: Klabenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-148

Query Match 23.2%; Score 285; DB 7; Length 238;
Best Local Similarity 30.1%; Pred. No. 1.7e-18;
Matches 69; Conservative 47; Mismatches 104; Indels 10; Gaps 3;

QY 6 ILIIVDELVTRNLKSIPEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLL 65
DB 4 ILIIEDEKNLAFVLELQHENYETVENNGRKGLODALLAQDFAILLDMLPDLNGLEI 63
QY 66 ARELREQANVALMFTTGRNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSRT--- 121
DB 64 ARVRQVKTTPIMLTARDSDVIDRVSGLDHGADDIVRPPALEELLARLRAVLRVKIEK 123
QY 122 -MNLGTVSEERSVSYKPNKGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKIQ 180
DB 124 DASKVTVAQK---IVKFKDLTIETANRIVHRGDKAIDLTKREYNLLMTLIENKNVV 179
QY 181 SPAELKKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIATIHGEYR 230
DB 180 SRDQLLNKIGWPESTIETNVVRYLYLNKIDA-PGQPSYIKTVRGTYG 228

RESULT 3
US-10-467-657-2276
; Sequence 2276, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 2276
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2276

Query Match 22.3%; Score 274; DB 6; Length 225;
Best Local Similarity 30.1%; Pred. No. 1.5e-17;
Matches 71; Conservative 47; Mismatches 92; Indels 26; Gaps 6;

QY 6 ILIIVDELVTRNLKSIPEAGYDVFEATDG-AEMHQILS-EYDINLVIMDINLPKNGL 63
DB 4 VLLVDDALLTELLTEYLSAEGLNVRSPDGEAGVQEILSGQYD--VVLDSPMPKNGL 61

QY 64 LLARELREQANVALMFTTGRNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSRTMN 123
DB 62 DVLKNVRARSTVPIIMLTAKGDDIDRIIGLEWGADDYVFKPCTPRELLARINAILURRAQH 121
QY 124 LGTVSEERSVSYKPNKGWELDINSRLSGPDGSGQ-----YKLPSEFRAMLHFCENP 176
DB 122 SGEQNNAFNSISV-----SDVLYIPAKRQASVKDMPLELTSFEFNLLEVLMRHA 170
QY 177 GKIOSRAELKKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIATIHGEYR 232
DB 171 GQVWSKETLSVEALDRKLAKFDRSIDVHISIRHKLGD-----SLIQTVRGGLGYLF 222

RESULT 4
US-10-793-626-3266
; Sequence 3266, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3266
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3266

Query Match 21.3%; Score 262; DB 6; Length 224;
Best Local Similarity 28.9%; Pred. No. 1.8e-16;
Matches 67; Conservative 45; Mismatches 96; Indels 24; Gaps 5;

QY 7 LIVDELVTRNLKSIPEAGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGLLLA 66
DB 5 LIVDDKKLQYVSSHLERESIQTHFTTSGEASLDLFLENKNVDIAIVDIMKSGMGDFELC 64
QY 67 RELREQANVALMFTTGRNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSR----- 120
DB 65 QTLKDDYHIPVIMLTARDALSDKERAFLSGTDYVTKPFVKELAFRIKAVLRRYQINAD 124
QY 121 -TMNLGTVSEERSVSYKPNKGWELDINSRLSGPDGEQYKLPSEFRAMLHFCENPGKI 179
DB 125 NELQGNLIIN-----QSY-----MEITVGSKTM-----NLPNKEFQLFLASNPKH 168
QY 180 QSRAEKKMTGRELKPHDRTVDVTIRIRKHFESTPDTPEIATIHGEYR 231
DB 169 FTRODIIIEKINGFDYEGDDRTVDVHIEKRLQRLSKLKSSVS-IQTVRGQYR 219

RESULT 5
US-10-793-626-1424
; Sequence 1424, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1424

```
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1424

Query Match      16.6%; Score 204; DB 6; Length 166;
Best Local Similarity 28.9%; Pred. No. 1.8e-11;
Matches 52; Conservative 33; Mismatches 71; Indels 24; Gaps 4;

Qy 57 LPKNGKLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARN 116
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 LPNGLGICQIRQKTTPIIIITAKSETYDKVAGLDYGADDYIVKPFDELLARIRA 61

Qy 117 LLSRTMNLGTSEERSRSVSYKFNCGWELDNRSRL-----IGPDGEQYKLPSEFRAML 170
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 VLRR-----QPKDQV-----LDINGIIDKDAFKVTNNGHQLEUTKTEYDILY 104

Qy 171 HFCENPGKIQRGAELKKMTGRELKPHDRTVDVTIRIRKHFESPTDTPPEIATIHGEGY 230
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
105 VLAENRNHVMQREQILDHWGYNSEVETNVDDVYIRYLNRNKLKP-NKEKSIETVRGVGY 163

RESULT 6
US-10-793-626-2788
; Sequence 2788, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2788
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2788

Query Match      16.6%; Score 204; DB 6; Length 166;
Best Local Similarity 28.9%; Pred. No. 1.8e-11;
Matches 52; Conservative 33; Mismatches 71; Indels 24; Gaps 4;

Qy 57 LPKNGKLLARELRQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARN 116
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 LPNGLGICQIRQKTTPIIIITAKSETYDKVAGLDYGADDYIVKPFDELLARIRA 61

Qy 117 LLSRTMNLGTSEERSRSVSYKFNCGWELDNRSRL-----IGPDGEQYKLPSEFRAML 170
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 VLRR-----QPKDQV-----LDINGIIDKDAFKVTNNGHQLEUTKTEYDILY 104

Qy 171 HFCENPGKIQRGAELKKMTGRELKPHDRTVDVTIRIRKHFESPTDTPPEIATIHGEGY 230
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
105 VLAENRNHVMQREQILDHWGYNSEVETNVDDVYIRYLNRNKLKP-NKEKSIETVRGVGY 163

RESULT 7
US-10-873-528-54
; Sequence 54, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
```

```
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 54
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-54

Query Match      16.1%; Score 198.5; DB 6; Length 185;
Best Local Similarity 30.3%; Pred. No. 6.4e-11;
Matches 47; Conservative 28; Mismatches 63; Indels 17; Gaps 2;

Qy 3 TPHILIVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNG 62
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 TKQVLLVDDERHILKLLDYHLKSGFSTQLVNTRKALALAEPTFPDFILLDMLPQLDGG 61

Qy 63 LLARELREQ-ANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLLSRT 121
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 MEVCKRLRAKGVKTPIMVSAKSDEFKVLALALEIGADDYITKPFSPRELLARVAVLRT 121

Qy 122 MN-----LGTVSEERSVESYKFN 140
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 KGEQEGDSDNADDSWLFGLTKVYPERHEVYKAN 156

RESULT 8
US-11-098-686-11387
; Sequence 11387, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11387
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-11387

Query Match      14.2%; Score 175; DB 7; Length 122;
Best Local Similarity 31.4%; Pred. No. 4.5e-09;
Matches 37; Conservative 29; Mismatches 48; Indels 4; Gaps 2;

Qy 5 HILVDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLPKNGKL 63
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 HILVDDSKTVRNLMSPFVKSQGVTTAEQDLGLEKLYSVKVDKVLIVSINMRMDGF 63

Qy 64 LLARELREQ-ANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRLTIRARNLL 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 TFIKSVREQDLYRNPIIILSTEKEDEDIKGMKFGANLYMVKPAQPDPTLVNRNKL 121

RESULT 9
US-10-467-657-1982
```


Sequence 1982, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SpA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 1982
LENGTH: 188
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1982

Query Match 12.5%; Score 154; DB 6; Length 188;
Best Local Similarity 29.5%; Pred. No. 6.2e-07;
Matches 38; Conservative 26; Mismatches 45; Indels 20; Gaps 2;

Qy 27 GYDVFEATDGAEMHQILSEYDINLVIMDINLPGKNGLLARELREQ-ANVALMFLTGSDN 85
Db 9 GYADVWVKNQAAVAAAAQPYDMLDLGLPGDGLDVLSEIRAACTVPVLIVTARD 68
Qy 86 EVDKILGLEIGADDDYITKPFNPRLTIRARNLLSR-----TWNLGT 126
Db 69 LYSRLNGLDGADDYIVKPPDMAEPKARMAVLRRSGGQAQCLNGLSLNPATVQVEI 128
Qy 127 VSEERRSVE 135
Db 129 IAEGRGGIE 137

RESULT 10
US-10-467-657-8052
Sequence 8052, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SpA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 8052
LENGTH: 425
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8052

Query Match 11.9%; Score 146.5; DB 6; Length 425;
Best Local Similarity 26.5%; Pred. No. 9e-06;
Matches 40; Conservative 34; Mismatches 62; Indels 15; Gaps 2;

Qy 1 MQTPHILIVDELVTRNTLKSIFEAAGYDVFEATDGAEMHQILSEYDINLVIMDINLPCK 60
Db 1 MRSSDILIVDVEGIRDLSEILQDEGYSVALAENAEARKLRHQARPANVLIDIMPD 60
Qy 61 NGLLARELRE--QANVALMFLTGSDNEVDKILGLEIGADDDYITKPFNPRLTIRARNLL 118

Sequence 1982, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SpA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 1982
LENGTH: 188
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1982

Query Match 12.5%; Score 154; DB 6; Length 188;
Best Local Similarity 29.5%; Pred. No. 6.2e-07;
Matches 38; Conservative 26; Mismatches 45; Indels 20; Gaps 2;

Qy 27 GYDVFEATDGAEMHQILSEYDINLVIMDINLPGKNGLLARELREQ-ANVALMFLTGSDN 85
Db 9 GYADVWVKNQAAVAAAAQPYDMLDLGLPGDGLDVLSEIRAACTVPVLIVTARD 68
Qy 86 EVDKILGLEIGADDDYITKPFNPRLTIRARNLLSR-----TWNLGT 126
Db 69 LYSRLNGLDGADDYIVKPPDMAEPKARMAVLRRSGGQAQCLNGLSLNPATVQVEI 128
Qy 127 VSEERRSVE 135
Db 129 IAEGRGGIE 137

RESULT 10
US-10-467-657-8052
Sequence 8052, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SpA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 8052
LENGTH: 425
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8052

Query Match 11.9%; Score 146.5; DB 6; Length 425;
Best Local Similarity 26.5%; Pred. No. 9e-06;
Matches 40; Conservative 34; Mismatches 62; Indels 15; Gaps 2;

Qy 1 MQTPHILIVDELVTRNTLKSIFEAAGYDVFEATDGAEMHQILSEYDINLVIMDINLPCK 60
Db 1 MRSSDILIVDVEGIRDLSEILQDEGYSVALAENAEARKLRHQARPANVLIDIMPD 60
Qy 61 NGLLARELRE--QANVALMFLTGSDNEVDKILGLEIGADDDYITKPFNPRLTIRARNLL 118

THIS PAGE BLANK (USPTO)